Genomic Diversity within the "Actinomyces naeslundii" Group

Mughal, Sadaf Rasheed

Awarding institution:
King's College London

Genomic Diversity within the

"*Actinomyces naeslundii*"

Group

Thesis submitted in accordance with the requirements of the

King's College London for the degree of Doctor in Philosophy

by

Sadaf Rasheed Mughal

2014

Unless otherwise stated, this thesis is the original work of the author.

Sadaf Rasheed Mughal

Published Paper:

Beighton, D., Sadaf Rasheed Mughal and Thuy Do (2013). Bacterial Catabolism of Salivary Substrates. Oral Microbial Ecology: Current research and New perspective, 37-46.

ABSTRACT

The aim of the study was to determine whether sub-clusters identified within *A. oris* and *A. naeslundii* by Multi locus sequence typing (MLST) of the concatenated sequences of 7 housekeeping genes should be classified as subspecies or distinct species. Determination of whole genome sequences of selected strains of *A. oris* and *A. naeslundii*, using Illumina Genome Analyzer IIxe and Roche 454 methods was initiated using paired-end sequencing and single reads respectively. The sequences obtained were aligned using CLC software and annotated using RAST. In preliminary gene-by-gene, analysis a putative Lacto-*N*-Biose [LNB] gene operon was identified which may mediate growth of *A. oris* on LNB generated *in vivo* from the degradation of salivary mucins. The operon were found in all *A. oris* strains (n=19) but in 18 *A. naeslundii* strains the operon was dysfunctional with extensive, varied gene loss apparent. Mutants of *Actinomyces oris*-MG1 were produced with the LNB phosphorylase gene inactivated. All *A. oris* strains grew on LNB (increase in OD= 0.213 $\pm$ 0.07 and final pH= 4.99) while none of the *A. naeslundii* strains (increase in OD=0.09 $\pm$ 0.05 and final pH=6.3; both $p<0.01$) or the *A. oris* mutant did not grow significantly in the presence of LNB. This finding indicated that *A. oris* strains possess a functional LNB operon and so are better able to utilize LNB generated by the degradation of salivary O-linked glycans. This is the first demonstration of LNB utilization by a member of the oral microbiome and may explain the greater prevalence of *A. oris* compared to *A. naeslundii* in oral biofilms. Further the Reverse Transcriptase-PCR (RT-PCR) was carried out to demonstrate the up-regulation of LNBP gene when *A. oris*-MG1 was grown in LNB. Apart from bio-chemical characterization, comparative genomic analysis was carried out on whole genomic sequences. Digital DNA-DNA homology (DDH) values were obtained using in-silico genome-to-genome comparison and this analysis delivered results consistent with previous traditional classification. Evolutionary analysis using ClonalFrame was also observed. The mutation and recombination events were compared using chi-square test among *A. oris* and *A. naeslundii* isolates. *A. oris* mutation and recombination event were significantly higher than *A. naeslundii* showing the diversity of *A. oris* strains in the oral cavity. Seven house keeping genes of publically available strains were included in previous MLST analysis and it was investigated that k20, MG1, c505, OT175 clustered in *A. oris* group of isolates while OT171, OT170 and *A. johnsonii* appeared as separate branches in MLST analysis and same was observed in core genome phylogenetic tree. These findings suggest that *A. oris* forms six distinct groups based on Neighbour-Joining tree analysis, core genome alignment and DDH analysis while *A. naeslundii* forms a single compact cluster.

TABLE OF CONTENTS

TABLE OF FIGURE

TABLE OF TABLES

ABBREVIATIONS

β-NAG	β-N-acetyl-glucosaminidase
CAMP	Synergistic hemolysis test
DDH	DNA-DNA Hybridization
DNA	Deoxyribo Nucleic Acid
h	hour
HMO	Human Milk Oligosaccharides
HOMD	Human Oral Microbiome Database
HI	Heart Infusion broth
HSP	High-scoring segment pairs
IGA	Illumina Genome Analyzer
IgA	Immunoglobulin A
LNB	Lacto-N-Biose
Mb	Million bases
min	minute
ml	Milli lire
µl	Micro-litre
MLST	Multi Locus Sequence Typing
MUMs	Maximally unique Matches
NGS	Next Generation Sequencing
pM	pico Molar
PMN	polymorphonuclear leukocytes
QF	Qubit Fluorometer
RBC	red blood cells
RE	Recombination event

rpm	revolution per minute
RT-PCR	Reverse Transcriptase-PCR
spp.	species
v/v	volume by volume
w/v	weight by volum

ACKNOWLEDGEMENTS

This project is by far the most significant accomplishment in my life and it would not have been possible without people who supported me and believed in my caliber. This project was kindly funded by Higher Education Commission of Pakistan and King's College London.

I would like to extend my gratitude and sincere thanks to my honorable supervisor **Prof. David Beighton.** He is not only a great supervisor with deep vision but also most importantly a kind person. I sincerely thank for his exemplary guidance and encouragement. His trust and support inspired me in the most important moments of making right decisions and i am glad to work under his supervision. I am also thankful to my second supervisor **Dr. David Radford** for helping me to proof read the thesis and finally for viva preparation.

I express my sincere thanks to the people in Department of Microbiology, particularly Dr. Thuy Do, Dr. Sadia Niazi, Mr. Steven Gilbert, Dr. Douglas Clark, Mr. Ahmed Hawas, Ms Lynnette for the encouragement, help, and for the good time i had during the project. I am also thankful to Dr. Efterpi Papouli and Ms. Muddasar Mirza of Biomedical Research Centre, King's College London for allowing me to avail the cutting edge facilities of High-throughput sequencing of Illumina Genome Analyzer and Dr. Tony Brooks of Univeristy College London, for helping me to analyze DNA samples using Roche 454 Genome Sequencer. I also would like thank Dr. Xavier Didelot of Imperial College London for helping me to do ClonalFrame Analysis of whole genome sequence data. Lastly i would like to say thanks to Dr. Hung Ton-That of The University of Texas Health Science Centre at Houston, Medical School, Department of Microbiology and Molecular Genetics, USA for allowing me to visit their lab and providing me the facilities related to generation of knock-out mutant strains of *Actinomyces* and also to Dr. Wu-Chengang (post-doc colleague in Houston University, Texas, USA) for helping me to successfully implement the allelic exchange technique.

I would like to thank my Father (Abdul Rasheed Mughal), Mother (Shakeela Rasheed Mughal), Grandparents, Brothers (Sohail, Rohail and Babar), Sisters (Sobia, Summreen, Rida and Nida) and most importantly my husband (Muhammad Akhtar Khan) and sons (Muhammad Ashar Yousaf and Muhammad Ayan) for the prayers, patience, support and tolerance to gave me throughout all stages of PhD project and life.

Chapter 1 INTRODUCTION

1.1 The Oral Micro-flora

The micro-organisms found in the oral cavity of humans may be known collectively by the terms oral microflora, oral microbiota or oral microbiome. The oral microbiome, a term recently used to describe micro-organisms found in the oral cavity was introduced by investigators of the Human Oral Microbiome Project (Dewhirst *et al.*, 2010). The human microbiome was defined as the "ecological community of commensal, symbiotic, and pathogenic micro-organisms that literally share our body space" (Lederberg, 2001). The investigators of the Human Oral Microbiome Database (HOMD) project believe that the study of human health and disease is not possible without clear understanding of the identification and phylogenetics of the oral microbiota. Both culture and culture-independent methods have been used to uncover the abundance of oral bacteria for many years. There are 280 bacterial species in the oral cavity which have been identified by culture methods. There are 700 common oral taxa (phylotypes) which have been identified in the human oral microbiome using culture-independent methods such as 16S rRNA gene based cloning and sequencing studies as defined and accessible through the HOMD (Dewhirst *et al.*, 2010; Paster *et al.*, 2006). The representatives of phyla which have been found in the oral cavity include *Actinobacteria, Firmicutes, Fusobacteria, Spirochaetes, Proteobacteria,* and *Bacteroidetes* (Dewhirst *et al.*, 2010). The community of micro-organisms residing in the oral cavity varies between individuals and is unique to each individual (Aas *et al.*, 2005; Bik *et al.*, 2010)

The contents of this chapter include an overview of oral microbiota that plays an integral role in initiation and progression of dental plaque formation. In addition, it also highlights the significance to emphasize the diverse nature of this micro-flora, which was investigated previously by old cultivation methods/molecular sequencing methods and is now supplemented with the latest cutting edge next generation sequencing (NGS) techniques. In the current study, NGS techniques were employed to delineate the selected *Actinomyces* isolates and to investigate in detail, as a useful and rapid tool, to study the taxonomy of a group of "Actinomyces-naeslundii". The investigation of genomic diversity of complete genome sequenced strains of *Actinomyces* species

comprise the major part of the thesis and are extensively investigated using bioinformatic approaches.

1.1.1 Oral Cavity Act As A Complex Habitat For Bacterial Colonization

The oral cavity is unique and complex due to its ecological characteristics. The oral cavity contains various surfaces that create diverse ecological niches and each surface harbors its own set of micro-organisms (Marsh *et al.*, 2010; Nobbs *et al.*, 2011). Tongue, cheeks, palate and lips (mucosal surfaces), teeth and gingival crevice have their own unique microbial colonization patterns (Aas *et al.*, 2005). The presence of various habitats on each surface further adds to the complex nature of mouth (Bik *et al.*, 2010; Dethlefsen *et al.*, 2007). In addition to the natural complexities of the mouth the internal properties of the oral environment may also be changing on a regular or irregular basis. These regular/irregular changes may be mediated by the consumption of different foods including fermentable carbohydrates, hot and cold drinks, oral hygiene practices and salivary flow rate.

1.1.2 Bacterial Population In The Oral Cavity

The human body consists of only 10% of cells whose origin is mammalian out of the total 10^{14} cells which make up the commensal microbial population (Marsh *et al.*, 2010; Sanders, 1984; Wilson, 2005). In a healthy human body, saliva contains about 10^{8} cells/ml and gingival crevicular fluid consist of approximately 10^{11} cells/ml of bacterial population (Evaldson *et al.*, 1982). Bacteria in saliva are derived from the biofilms on the dentition and mucosal surfaces (Aas *et al.*, 2005; Marsh *et al.*, 2010).

1.2 Oral Bacteria And Dental Plaque / Biofilms

Oral microbial communities on the dentition are known by the term dental plaque. The general understanding is that microbiological studies start from the invention of the primitive microscope by Antonie van Leeuwenhoek in the 17th century. He was the first to observe the "animalcules" in the plaque formed on his own teeth (Aparna & Yadav, 2008; Donlan & Costerton, 2002). W.D. Miller later on introduced the term to describe the mechanism of caries (chemicoparasitic) and he was of the opinion that dental plaque was made of micro-organisms (Kuramitsu *et al.*, 2007). The micro-organisms live in the form of organised multi-species communities on the surface of the teeth (Palmer *et al.*, 2011).

1.2.1 Oral Biofilm Formation

The theory of biofilm development was first introduced by Costerton (1978) which suggested the mechanism used by bacteria to stick to the living or non-living surface. The tenacious adherent deposits formed on teeth may be called dental plaque or oral biofilm. The stages of dental plaque/oral biofilm formation include five distinct phases as shown in Figure 1.1. The phases of dental plaque formation include pellicle formation, early colonization by pioneer bacteria, secondary colonization and formation of mature biofilm or dental plaque (Figure 1.2)

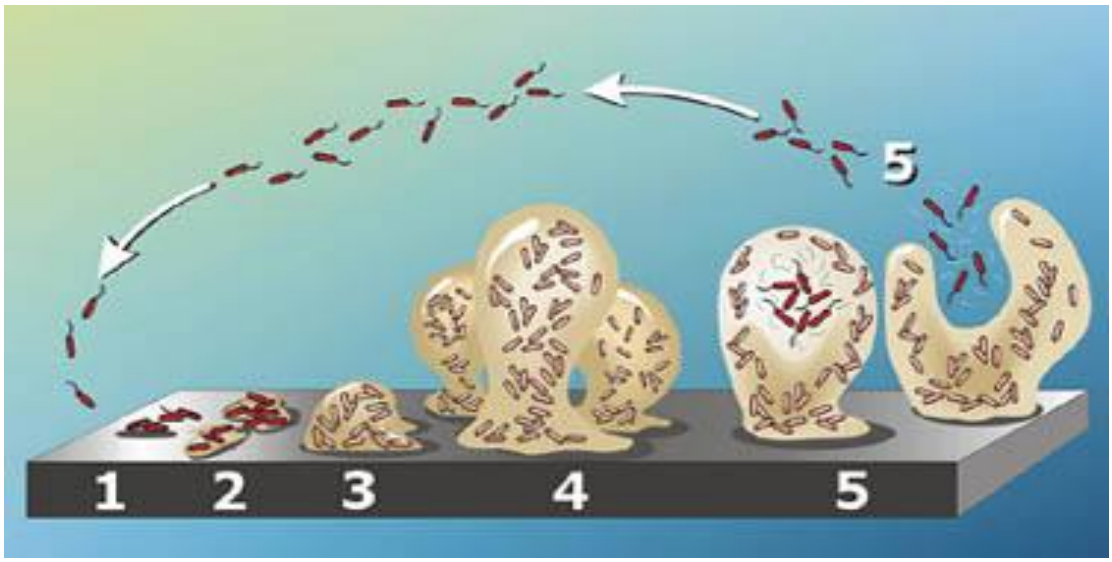


Figure 1.1: Stages during biofilm development.

*Stage1: Planktonic cells adhere to the surface, Stage 2: Exopolysaccharides secreted by these planktonic bacterial cells, Stage 3: Development of early architecture of biofilm, Stage 4: Biofilm architecture matures and during stage 5 planktonic cells starts dispersing from the mature biofilms. This figure is taken from Stoodly *et al*., (2002)

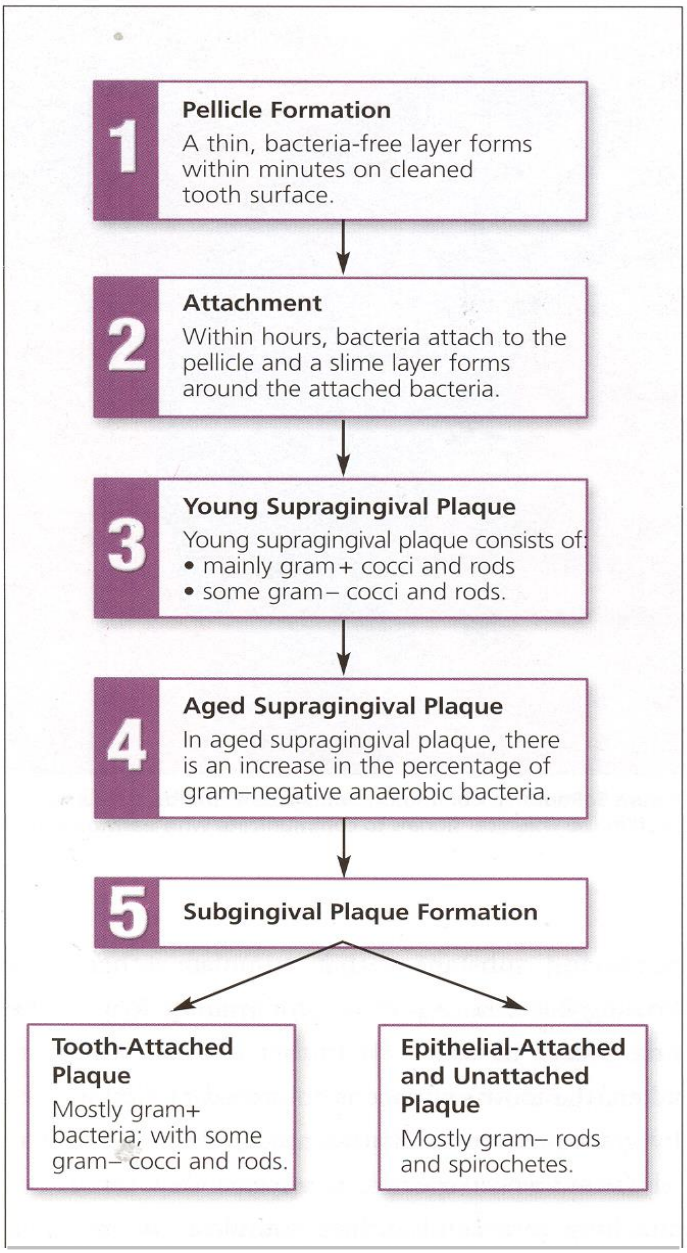


Figure 1.2: Phases of Subgingival plaque formation (Nield-Gehrig, 2003)

1.2.1.1 Acquired Pellicle Formation

A tooth, which was freshly polished, would remain clean in the mouth for only few minutes and after that, the tooth surface would be exposed to saliva to form the acquired salivary pellicle. The pellicle is a bacteria-free thin layer formed on the tooth surface. It consists mainly of salivary proteins but few non-salivary derived proteins, glycoprotein, lipids and carbohydrates are also present (Lendenmann *et al*., 2000). This layer serves a role of lubrication, protecting against demineralization and promoting remineralization and it mediates the composition of early microbial flora that colonizes during the first few hours of plaque formation. The pellicle formation is usually completed within the first 2 hours of protein accumulation and desorption (Lendenmann *et al*., 2000; Masson *et al*., 2013; Siqueira *et al*., 2012). Salivary constituents e.g. albumin, glycoproteins, mucins, cell debris and proline rich proteins etc provide variety of receptors, which can be recognized by bacteria, and bind on the surface of the tooth thus forming a thin pellicle layer. The pellicle structure is amorphous, acellular, organic and invisible until stained with an extrinsic dye.

1.2.1.2 Early Bacterial Colonization of the Pellicle By Pioneer Species

In the second phase of biofilm formation, the early colonizers recognize receptors on the pellicle surface within the first four hours and adhere to the pellicle forming the first bacterial layer (Lie, 1977; Lie, 1979). Salivary proteins are the primary source of nutrients for micro-organisms (Bradshaw *et al.*, 1994; Marsh *et al.*, 2010) because nutrients obtained from the diet are rapidly cleared from the mouth (Beighton, 2013; Keene *et al.*, 1966). Saliva mediates the first interactions between bacteria to adhere to the pellicle and can be reversed due to its weak interaction to the cell surface (Kolenbrander *et al.*, 2002). Bacteria bind to pellicle irrespective of the host diet (Marsh *et al.*, 2010). The Gram-positive cocci, rods and small number of Gram-negative cocci are among the first micro-organisms that are laid down on the surface of the teeth within

the first hour. The species that predominate during the early phase of dental plaque formation are a specific subset of micro-organisms comprising of *Streptococcus, Actinomyces, Veillonella, Prevotella* and *Neisseria* species (Diaz *et al.,* 2006; Li *et al.,* 2004; Marsh *et al.,* 2009; Nyvad & Kilian, 1987; Nyvad & Kilian, 1990; Periasamy & Kolenbrander, 2010). Streptococci constitute 60-80% of dental plaque formed during the first 4-8 h of pellicle attachment (Dige *et al.,* 2009; Jakubovics & Kolenbrander, 2010). *Streptococcus oralis, Streptococcus mitis, Streptococcus sanguinis* were among the most common colonizing species on tooth surface. This is followed by colonization by *Actinomyces*. The composition of bacteria in the mouth varies at different oral sites (Papaioannou *et al.,* 2009; Sachdeo *et al.,* 2008). Within a few hours extensive microbial growth may be observed (Jakubovics & Kolenbrander, 2010; Palmer *et al.,* 2006).

1.2.1.3 Outgrowth Of Pioneer Species On Tooth Pellicle

The attachment of early colonizers on the tooth surface not only changes the tooth surface but a new surface-attached phenotype develops with unique metabolic activity containing surface properties which enhances the attachment of other bacteria to colonize (Davey & Costerton, 2006).

1.2.1.4 Secondary Colonization

There are certain bacteria, which produce sticky extracellular substances from salivary proteins and polysaccharides. These bacteria start synthesizing proteins and further enhance the attachment of bacteria which results in the increase in number of bacteria (Nobbs *et al.,* 2011). These sticky sugars are a contineous source of energy for the plaque bacteria. The bacteria finally reach to the decline phase of growth and form plaque on the surface of tooth (Figure 1.3 & 1.5). Gram-negative bacteria are among the secondary colonizers laid down on the tooth surface. *Prevotella intermedia* and

Capnocytophaga are among the few micro-organisms of secondary colonizers (Rickard *et al.*, 2003; Rosan & Lamont, 2000).

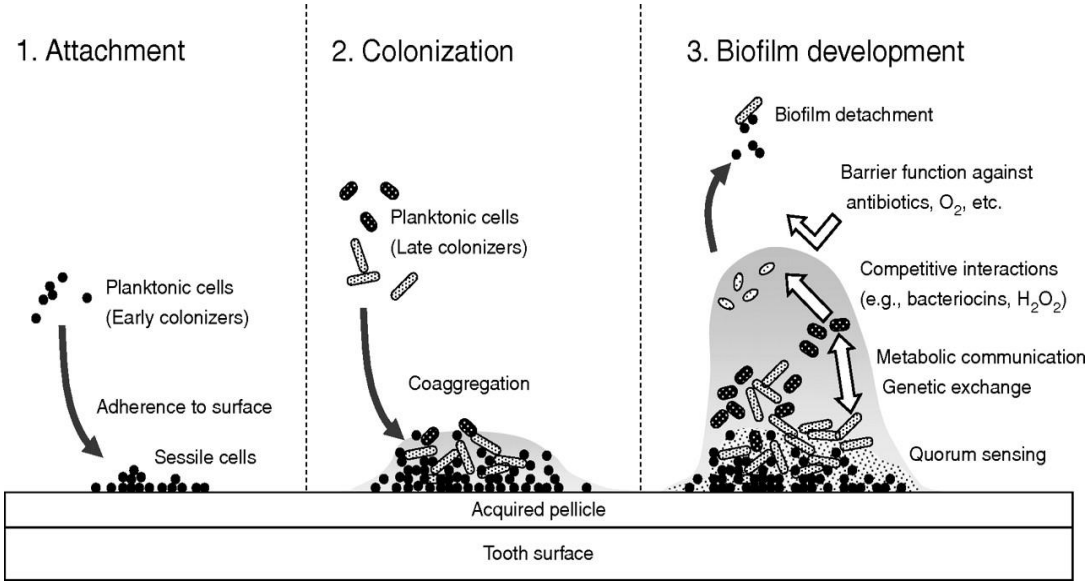


Figure 1.3: The illustration of biofilm development and the role of bacterial interactions (Hojo *et al.*, 2009)

1.2.2 Structure Of Biofilm

The dental plaque was recently considered as a biofilm due to the advent of confocal microscopy, which has revealed the resemblance of plaque to that of biofilms in other locations as shown in Figure 1.4. This unique resemblance of open structure features of dental plaque enable aerobic and anaerobic micro-organisms to maintain the flow of nutrients, enzymes and oxygen through the biofilm. The biofilms of dental plaque are the cause of oral diseases in the mouth (Nield-Gehrig, 2003).

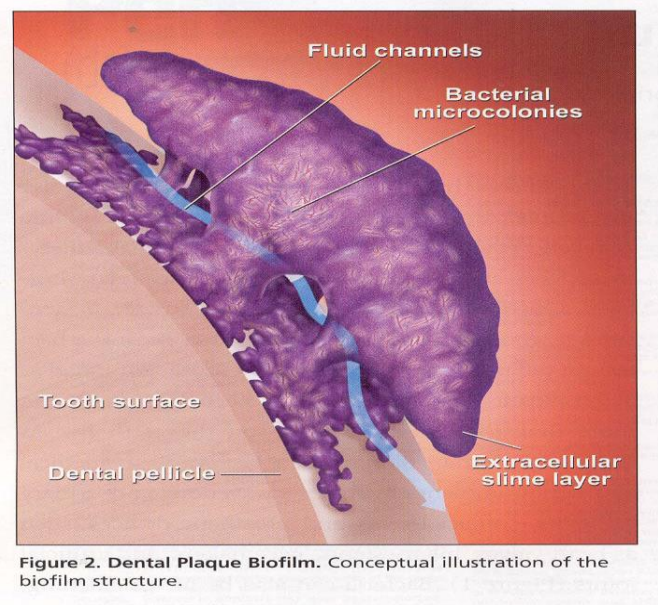


Figure 1.4: The structural representation of a biofilm (Nield-Gehrig, 2003)

1.2.3 Oral Bacterial Interactions For The Formation Of Mature Dental Plaque

Micro-organisms are accumulated in the mouth due to the warm and moist environment of the oral cavity. There are several factors which affect the growth of bacteria in the mouth e.g. narrow ranges of temperature, pH and oxidation-reduction potential (Marsh *et al.*, 2009; van Houte *et al.*, 1996). Oral bacterial interactions play a vital role in the development of mature dental plaque. These interactions can be of various types; metabolic exchange, exchange of genetic material, signalling among bacterial cells and the physical contact. The mechanism involved in the oral bacterial interactions with the host is discussed here briefly.

1.2.3.1 Co-Aggregation And Co-Adhesion

Co-aggregation or adhesion has an important role in the growth of mature biofilms (Bos *et al.*, 1996). Co-aggregation is the mechanism by which distant bacterial species interact and clump together in suspension and arrange themselves in a specific pattern for the formation of dental plaque (Gibbons & Nygaard, 1970; Kolenbrander *et al.*, 1983; Kolenbrander & Andersen, 1986; Mishra *et al.*, 2010). The early colonizers co-aggregate with themselves but not with late colonizers (Kolenbrander *et al.*, 1993; Kolenbrander & London, 1993; Whittaker *et al.*, 1996). Co-adhesion also has an important role for the formation of plaque (Figure 1.6), it involves the binding of the bacterial species to other species already adhered to a surface which further mediates the growth of the biofilm (Bos *et al.*, 1996). These early colonizers mainly constitute *Actinomyces* and *Streptococcus* species as discussed earlier (Fig 1.5). The binding of these micro-organisms is mediated by statherin or proline rich proteins with the fimbriae (Nobbs *et al.*, 2011). The direct contact between oral species of same or distant origin was reported by Gibbons and Nygard (1970). Other bacteria, which play a role to proliferate biofilm, include *Fusobacterium nucleatum.* It is a Gram-negative bacterium, the role of which in the context of supragingival biofilm formation is to co-aggregate with both early and late colonizers as shown in Figure 1.5 (Andersen *et al.*, 1998; Bradshaw *et al.*, 1998; Kolenbrander & Andersen, 1989; Periasamy & Kolenbrander, 2010). Kolenbrander *et al.*, (2006) has demonstrated over 100 pairwise co-aggregation interactions among oral plaque-forming bacterial communities.

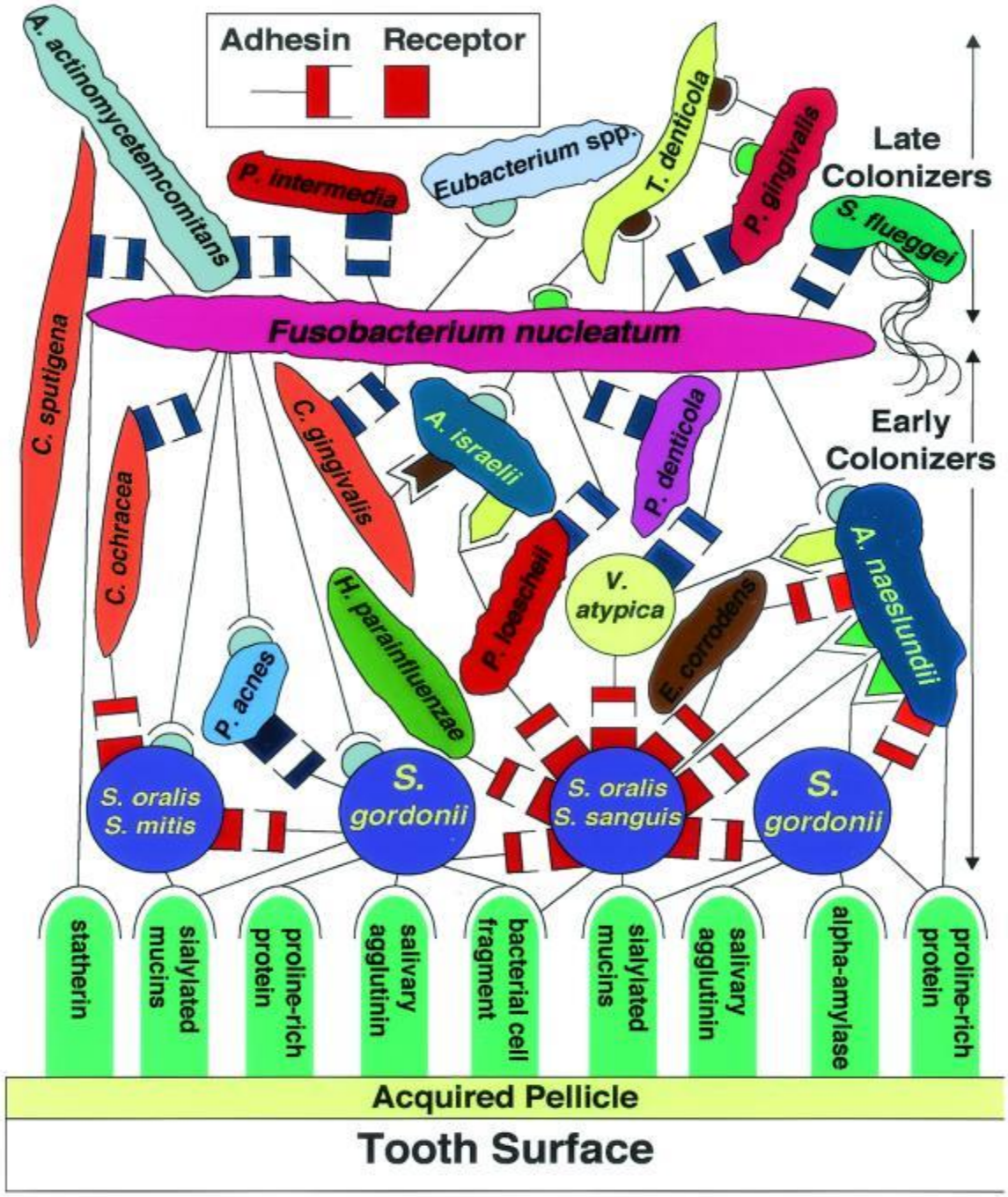


Figure 1.5: Plaque formation by early and late colonizers redrawn from Kolenbrander *et al.*, (2002).

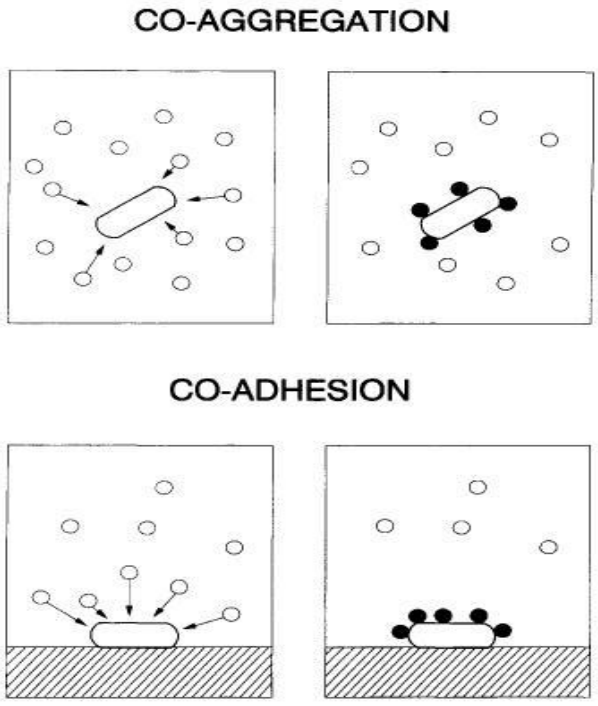


Figure 1.6: The illustrative mechanism of Co-aggregation and Co-adhesion (Cisar *et al.*, 1979)

1.2.3.2 Metabolic Bacterial Interactions In Dental Biofilms

The micro-organisms compete/cooperate with each other when present in the form of dental biofilms (Hojo *et al.*, 2009). There are different ways used to communicate by the micro-organisms in dental biofilms. These include exchange of metabolites, exchange of genetic material and quorum sensing (Kolenbrander *et al.*, 2006; Sedgley *et al.*, 2008).

Bacteria need to interact with salivary proteins to survive (Douglas, 1994). These interactions are aimed to provide nutrients, adhere to surfaces which results in aggregation or killing and therefore detachment of micro-organisms (Douglas, 1994). The main source of nutrients are available in the form of saliva, food containing sugars (easily fermentable carbohydrates), gingival crevicular fluid (GCF) and metabolic products excreted and secreted from micro-organisms (Wilson & Wilson, 2004). Similarly the excretion of a metabolite from one micro-organism was used by other micro-organisms (Cisar *et al.*, 1979; Kolenbrander *et al.*, 2002) and also the substrate breakdown had occurred by extracellular enzymatic activity of one micro-organism that was used as a source of nutrient for other micro-organisms in dental biofilm. Figure 1.7 shows metabolic bacterial interaction in dental biofilm and indicates that *Streptococcus*, *Lactobacillus* and *Actinomyces* species used sugars and saliva as a food source and secrete lactate which was utilized as a carbon source by *Veillonella* and *Propionibacterium* species (Kolenbrander *et al.*, 2006; Marcotte & Lavoie, 1998). Finally the menaquinone and its analogue was produced, which is useful for the growth of vitamin K-auxotrophic bacteria namely *Porphyromonas*, *Prevotella* and *Bifidobacterium* species (Hojo *et al.*, 2009; Marcotte & Lavoie, 1998).

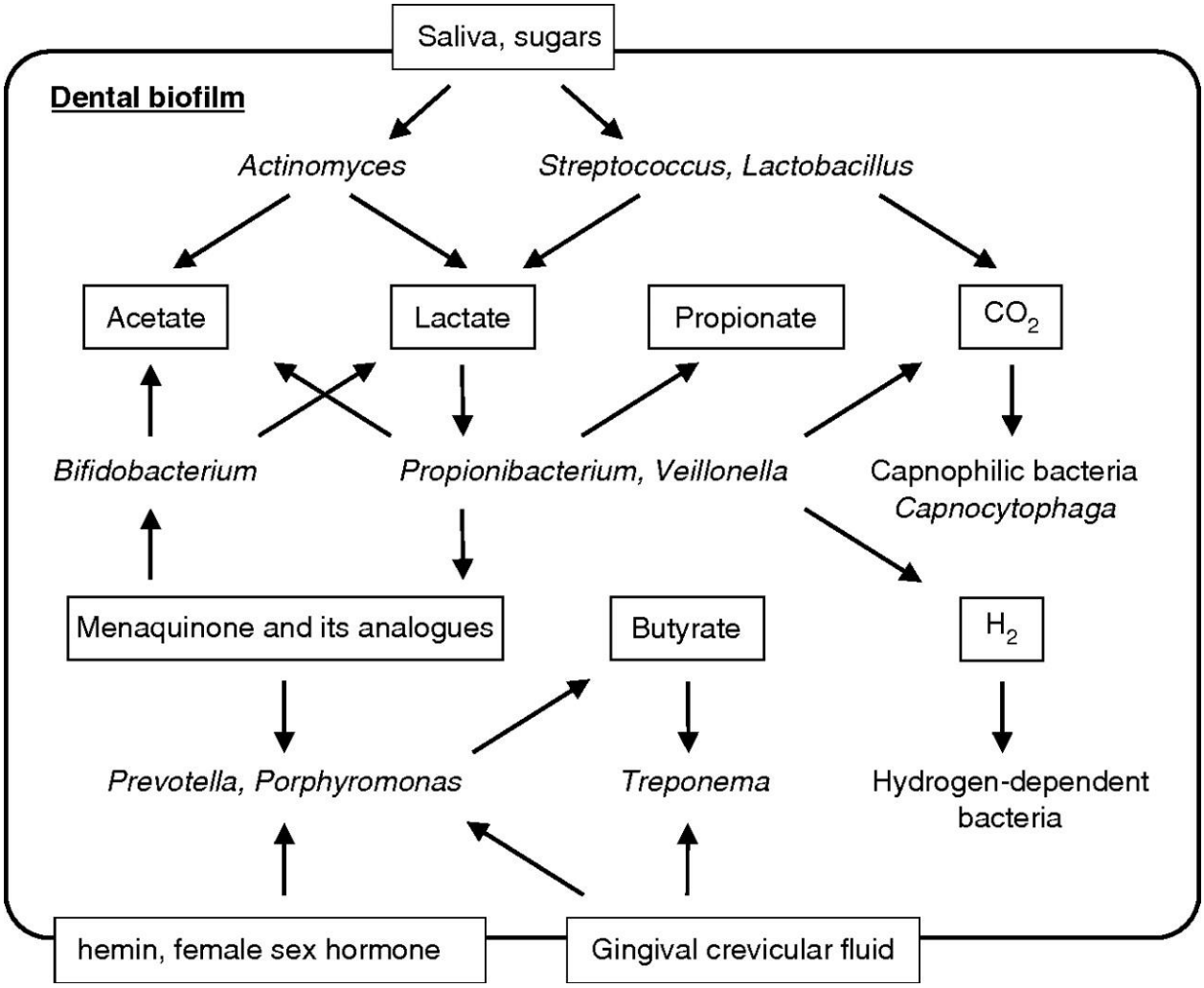


Figure 1.7: Illustration of postulated metabolic bacterial interaction in dental biofilm model (Hojo *et al.*, 2009)

Saliva is very important factor that contributes to the formation of plaque and the most abundant component in saliva is salivary amylase (Scannapieco *et al.*, 1989). Alpha-amylase is found in the acquired enamel pellicle. This suggested the role of alpha-amylase component of saliva in the adhesion of alpha-amylase-binding-bacteria. Alpha amylase helps to hydrolyze the starch components of food and this in turn provides glucose to bacteria present in plaque as their food source and contribute to the formation of more plaque on the tooth surface (Scannapieco *et al.*, 1989). In some early investigations, the bacteria that were found to involve abundantly to salivary alpha-

amylase adherence and utilization were oral streptococci (Gibbons *et al.*, 1983; Murray *et al.*, 1992; Scannapieco *et al.*, 1989; Whittaker *et al.*, 1996) and recent research has demonstrated the role of pili of oral *Streptococcus sanguinis* in binding to salivary amylase. The investigation was carried out by producing pili deficient mutant, lacking the genes encode for three pilus proteins named PilA, PilB, and PilC. The mutant was unable to bind to surface containing saliva. Further confocal microscopy observations suggested the lack of formation of the typical three-dimensional structure of plaque biofilm by the mutant and suggested the role of pili to promote the biofilm formation (Okahashi *et al.*, 2011; Scannapieco *et al.*, 1989).

Different studies have investigated whether the growth of micro-organisms is directly related to the ability of plaque bacteria to utilize endogenous nutrients from saliva especially glycoproteins. The salivary glycan degradation requires a set of enzymes called glycosidases and proteases (Bradshaw *et al.*, 1994). The glycosidases were known as sialidases, β-galactosidase, β-*N*-acetylglucosaminidase, β-*N*-acetylgalactosaminidase, α-β-mannosidase and α-fucosidase (Beighton & Whiley, 1990; Byers *et al.*, 2000). In one of the studies hog gastric mucin was used as a model glycoprotein and it played an important role in the growth of diverse bacterial communities (Bradshaw *et al.*, 1994). The mechanism involved in the metabolism of these molecules by which diversity of bacterial communities maintained was not fully understood but it was shown that glycoprotein catabolism was due to synergistic and combined action of many species having enzymatic activities to release nutrients from the glycoproteins and used as a source of energy (Bradshaw *et al.*, 1994). Another study on oral streptococci showed the ability to utilize mucin oligosaccharides as a source of energy. Digoxigenin-labelled lectins were used to trace the changes at the non-reducing end groups of oligosaccharide chain. *Streptococcus oralis* Ny 586 and *Streptococcus sanguis* Ny 584 were used in this study. It was found that core disaccharide Gal (1,3) GalNAc in O-glycans was degraded more effectively by *S. oralis* than *S. sanguinis.* Therefore the better growth of *S. oralis* on mucin was observed in this study (van der Hoeven & Camp, 1994). In another study MUC5B, a salivary mucin was cleaved by the concerted action of micro-organisms of freshly harvested supra-gingival plaque and dental plaque (Wickstrom *et al.*, 2009).

In addition to amylase and mucin, immunoglobulin A (IgA) is a very important constituent of saliva and 60 % of total salivary immunoglobulin (Ig) is IgA (Cheaib & Lussi, 2013; Mestecky, 1993). Salivary IgA has an important role in regulating plaque formation. In a study, salivary IgA levels were observed in patients with capability of slowly forming plaque shows higher IgA concentration as compared to patients with rapidly forming plaque. This finding suggested that salivary IgA has a very important role in the organisation of pioneer plaque formation (Zee *et al.*, 2001). The oral actinomyces and streptococcal lectins mediating adhesion was recognised by secretory Immunoglobulin A (S-IgA) (Ruhl *et al.*, 1996). S-IgA was reported to contain binding structure-Galβ1-3GalNAc, which is recognised by *A. naeslundii* (formerly known as genospecies I) but not for *A. oris* (*A. naeslundii* genospecies 2) as reported by Bratt *et al.*, (1999). One of the roles observed of sialidase produced by *Actinomyces* is used to desialylate IgA resulting in the proteolysis of cell (Ellen *et al.*, 1980; Mishra *et al.*, 2010).

The role of sialic acid appears to involve providing nutrition to oral biofilms. Sialidase activity was reported by many micro-organisms residing in human oral biofilms including *Actinomyces* spp. (Beighton & Whiley, 1990; Beighton *et al.*, 1991a; Braham & Moncla, 1992; Do *et al.*, 2008; Moncla & Braham, 1989). The research by Beighton and Smith (1986) showed that in fasting conditions the sialidase-producing bacterial proportion increased thus providing nutrition for the growth of other bacteria which are thought to be involved in biofilm formation. Another important finding reported was that early biofilm formation level reduced at night as compared to the day due to eating, which may suggest the differences in the availability of salivary nutrients between day and night (Dige, 2012).

Adherence of *Actinomyces oris* strains, previously known as *Actinomyces naeslundii* genospecies 2 (Do *et al.*, 2008; Henssge *et al.*, 2009), to glycans of salivary pellicle was mediated by sialidases. This exposes the galactose residue of O and N-linked glycans of salivary pellicle thus enhancing the adherence and colonization of specific bacteria to tooth forming oral biofilm (Costello *et al.*, 1979; Gibbons *et al.*, 1990). *A. oris* has two cell surface fimbriae (Figure 1.8). Type I fimbriae promote the adherence to salivary

proline rich proteins (Gibbons *et al.*, 1988; Mishra *et al.*, 2010), while type II fimbriae are involved in biofilm development and activation of host cells (Costello *et al.*, 1979; Ellen *et al.*, 1980; McIntire *et al.*, 1978). The host cells including red blood cells (RBC) and polymorphonuclear leukocytes (PMNs) contains receptors (Ruhl *et al.*, 1996; Stramberg & Karlsson, 1990). The activity of host cell receptors may be blocked by α-2-3 linked sialic acid. Sialidase, produced by *A. oris* containing type II fimbriae may mediate the attachment by desialylation of the host cell receptors. Fim P and Fim A was identified as a major subunits of type I and II fimbriae respectively by the pioneering work of Yeung and colleagues (Yeung & Ragsdale, 1997; Yeung *et al.*, 1998). In another study by Mishra (2010), the surprising results were obtained thus confirming the previous pioneer research work of Yeung by making a deletion mutant of *A. oris*- MG1 lacking type II fimbrial gene cluster. This was proved that the shaft of fimbriae, FimA, is solely responsible for the interaction of *Actinomyces* with *S. oralis* and in adhesion to erythrocytes, IgA, and more importantly in the formation of *Actinomyces* biofilms.

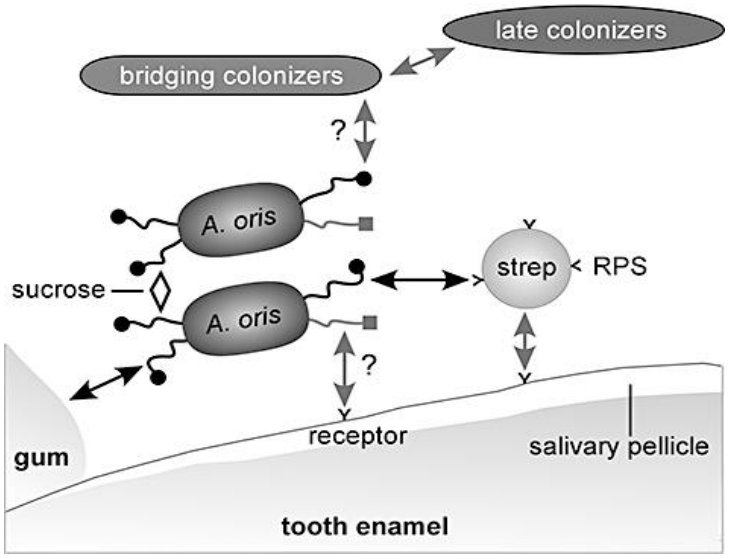


Figure 1.8: Importance of fimbriae in an adherence of *Actinomyces oris* to oral streptococci, tooth surface and host cells (Mishra *et al.*, 2010).

RPS: Receptor bearing-streptococci, fimbrial shaft FimP (grey), tip fimbrillin FimQ (grey square), fimbrial shaft FimA (black), tip fimbrillin FimB (black circle), unknown fimbrial components (?) and two way arrows represents interbacterial interaction and adherence to host cells.

Finally the micro-organisms organise themselves in the form of colonies. Socransky and Haffajee (2008) observed the colonization pattern of micro-organims in dental plaque biofilms. The clusters were created in terms of supra-gingival plaque formation based on nutritional and environmental differences and similarities among micro-organisms (Figure: 1.9). There is also an involvement of inter-species communication in dental plaque biofilms (Haffajee *et al.*, 2008).

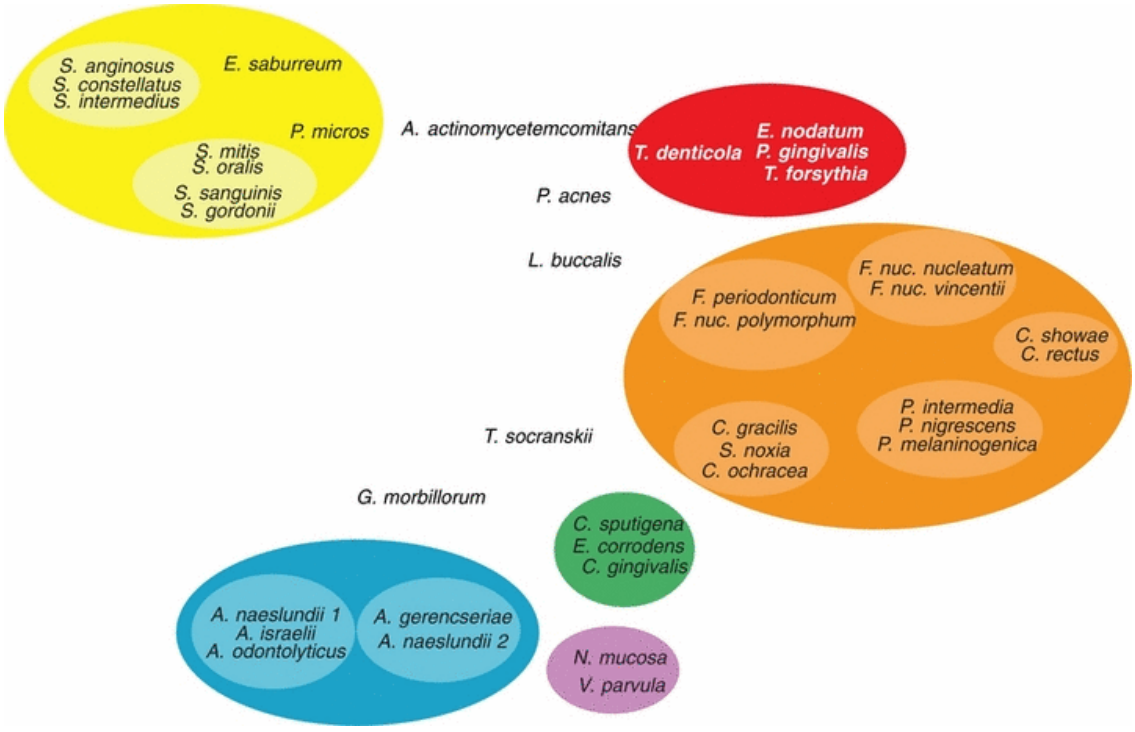


Figure 1.9: Relationship of species within and between microbial complexes in supragingival plaque samples (Haffajee *et al.*, 2008).

1.2.4 Pioneer Oral Bacteria And Their Role In Caries Development

Caries is a Latin word for decay and Dental caries defined as "...a localised chemical dissolution of the tooth surface caused by metabolic events taking place in the biofilm (dental plaque) covering the effected area..."(Fejerskov *et al*., 2008). This disease is a result of several factors. These factors include tooth structure, the micro-organisms involved and their load, the type of sugar consumption and their interaction with pioneer bacteria plays an important role. The bacteria, which were responsible to keep the dynamic stability on the surface of the tooth, were mainly non-mutans (streptococci and actinomyces). These bacteria on the surface of the plaque create an acidic environment through sugar metabolism causing demineralization. Once an acidic environment is established, these bacteria start increasing in number resulting in lesion formation. In this process, the surface texture of tooth enamel gets changed from smooth to rough and tooth dentine from hard to soft (Takahashi & Nyvad, 2008) . It has been shown that mutans streptococci (MS) are the major pathogens of dental caries and they have been found in caries lesions (Hamada & Slade, 1980; Loesche, 1986). Further, they are capable of producing glucan that does not dissolve in water and acts as a main mediator of attachment of bacteria to the tooth surface. Previous research revealed that MS are not always associated with dental caries; there is evidence of *Actinomyces* presence to initiate the caries process (Sansone *et al.*, 1993; van Houte *et al.*, 1996).

1.3 The Broad Classification Scheme

The broad classification scheme was universally accepted within the scientifi community. The order of classifying the organisms in descending order followed was Domain, Phylum, Class, Order, Family and Genus.

1.3.1 The Species Concept

Species is a word used in microbial taxonomy as a most fundamental unit. The boundaries of microbial species were considered "fuzzy" and have no universally accepted definition. The species concept consistently varied and was controversial in the past (Achtman & Wagner, 2008; Colwell & Russek-Cohen., 1995; Fraser *et al.*, 2009).

The most accepted and pragmatic species concept was described as a 'monophyletic and genomically coherent cluster of individual organisms that showed a high degree of overall similarity in many independent characteristics, and was diagnosable by a discriminative phenotypic property' (Rossello-Mora & Amann, 2001). This concept also called as a 'phylo-phenetic' species concept.

The concept of species for bacteria and archae was very difficult to describe, due to several reasons. The first major reason was that bacteria are asexual organisms and exchange their genetic material through either plasmid DNA, chromosomal DNA or bacteriophages but the frequency of exchange was still undetermined and perhaps the environmental selection pressure drives the diversification of bacteria. The technological limitation and poor understanding of functions were always hurdles to classify the phenotypic characteristics, which were very essential to describe the micro-organism at species level.

Another more accepted definition of prokaryotic species was 'a group of strains that show a high degree of overall similarity and differ considerably from related strain groups with respect to many independent characteristics', or 'a collection of strains showing a high

degree of overall similarity, compared to other, related groups of strains' (Colwell & Russek-Cohen., 1995).

The bacteria were always found in clumps of similar strains sharing many features with each other. The taxonomists agreed to implement a 'polyphasic approach' to classify bacteria which implies that two sources of information must be extensively studied e.g. genomic information and phenotypic information (Vandamme *et al.*, 1996). The genomic information was obtained directly from nucleic acids using available techniques of sequencing or indirectly from DNA-DNA similarity and G+C mol%. The information about micro-organisms available currently is based on DNA sequences, therefore it is difficult to classify the micro-organisms using their phenotypes due to non-availability of phenotypic information (Koeppel & Wu, 2013). A new strategy was adopted to overcome this challenge by deriving biologically and ecologically meaningful taxonomic units from DNA sequence information. The sequences were clustered based on the homology of 16S rRNA gene sequences and 97% identity cut off was used to delineate the species (Curtis *et al.*, 2002; Giovannoni & Stingl, 2005; Staley, 2006; Venter *et al.*, 2004).

Thus, it was obvious that bacteria were classified at species level by using sequence data from multiple or ideally seven core housekeeping loci. There were still no reproducible and successful methods, which could sharply define the unambiguous boundaries of species clusters. The boundaries of species was resolved with some fuzzy and indistinct forms but with multiple locus data, it was at least easy to develop the origin of species and relate it to observe the population genetic structures of bacteria (Hanage *et al.*, 2005). This proved to be extremely helpful for the taxonomist.

The bacterial classification at species level on a large scale was employed with the adaptation of molecular approaches especially DNA-DNA hybridization (Hanage *et al.*, 2005). DNA-DNA hybridization used as a gold standard to define bacterial species among the set of genomes who gave at least 70% similarity index (Hanage *et al.*, 2005; Pedros-Alio, 2006; Stackebrandt *et al.*, 2002). The inter-genomic distance was calculated using wet laboratory techniques to delineate the micro-organisms at species/sub-species level (Johnson *et al.*, 1990). The genomes of novel isolates were compared with the

genomes of known taxa to find out the similarity among genomes. However, this is time consuming and technically difficult. 16S rRNA gene has been used extensively as a tool to delineate the micro-organisms at species level especially for those micro-organisms that were still unculturable. 16S rRNA is a highly conserved gene and it was difficult to explore the similarity level among closely related sets of bacterial isolates. Recently, the advancement of high-throughput sequencing techniques led researchers to obtain high resolution genomes and DDH (DNA-DNA hybridization) similarity index was obtained and was used to describe the organisms at species level (Palys *et al.*, 1997; Palys *et al.*, 2000). In theory it is very complex to infer the overall similarity between the genomes of two strains but, due to progress in the field of gene sequencing, bioinformaticians are using in-silico genome-to-genome comparisons to replace the tedious wet laboratory DDH techniques (Auch *et al.*, 2010a; Auch *et al.*, 2010b; Henz *et al.*, 2005). There is a web service (http://ggdc.dsmz.de/) freely available to compare query and reference genomes.

Another accepted concept of species is that "species is a group of individuals where the observed lateral gene transfer within the group is much greater than the transfer between groups" (Dykhuizen, 2005). Nesbo *et al.* (2006) described the species as "Microbes…..do not form natural clusters to which the term "species" can be universally and sensibly applied". Ecotype and Genotype concepts are used side by side to describe the speciation process (Figure 1.10). Ecotype concept presented by Cohan (2002) with a role of recombination states that "An ecotype" (p 467) is a set of strains using the same or similar ecological resources, such that an adaptive mutant from within the ecotype out-competes to extinction all other strains of the same ecotype; an adaptive mutant does not, however, drive to extinction strains from other ecotypes". The Genetic concept of species shown in Figure 1.10 represents that genetic species A are the organism having similar genomes in which recombination occurs more frequently, while genetic species B are those whose genomes that are quite different and therefore recombination between them is not a common mechanism. Lateral transfer of genes shown as stars and circles in Figure 1.10, emerged into two species types called "Ecological species Y and Z".

Ecological species Y preferred to remain as single ecotypes while the other ecotype transformed into ecological species Z (Nesbo *et al.*, 2006).

The present species concept was emerged after successfully implementing the high-throughput techniques through which both genetic and phenotypic traits can be explored. The entire bacterial genome was a full source of genomic information. It was not possible to get whole genome sequencing data at large scale in the past but recently with the advent of the most powerful genome sequencing approaches, it was possible to employ the techniques readily and uncover the genetic and phenotypic information. The present study is part of this phenomenon to uncover the taxonomic status of a selected group of bacterial strains (*Actinomyces*) and is used in an attempt to assign correct naming at species/sub-species level.

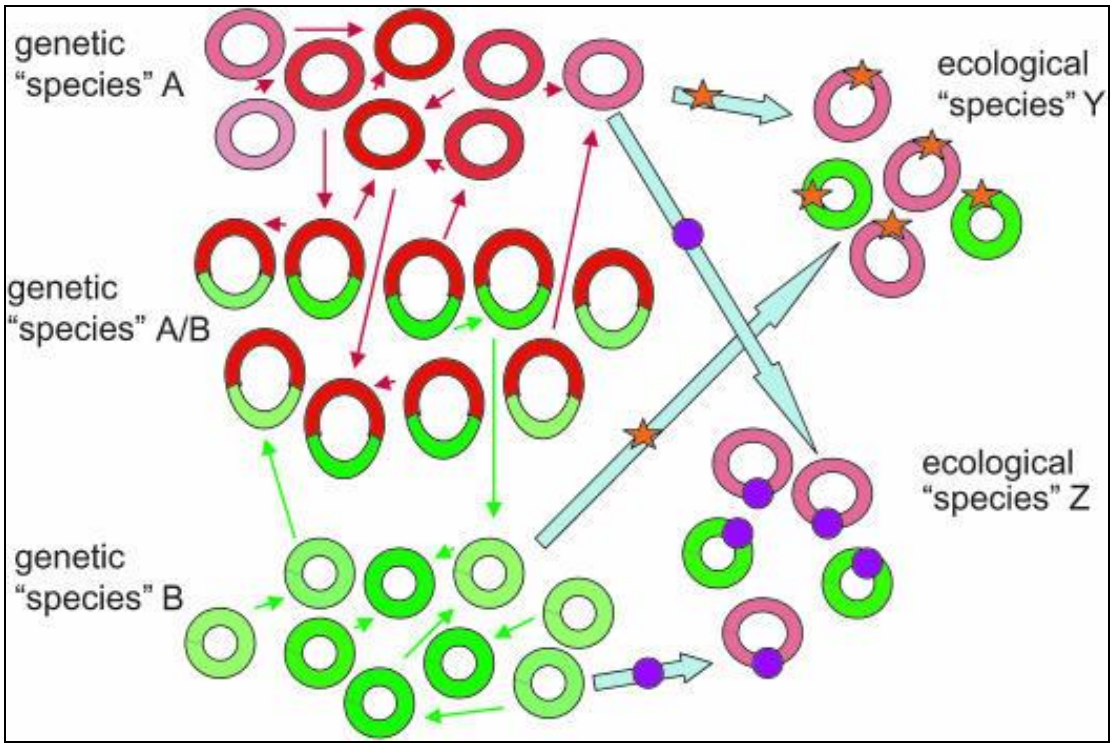


Figure 1.10: A simple model to represent the species concept emerged due to genetic and ecological factors (Nesbo *et al.*, 2006).

***Small arrows represent frequent recombination among closely related genomes, yellow stars and purple circles indicate lateral transfer of genes**.

1.4 Microbial Diveristy In the Oral Cavity

The bacterial communities are extremely diverse in the oral cavity (Aas *et al.*, 2005; Nasidze *et al.*, 2009; Segata *et al.*, 2012). The diversity of bacteria is due to several factors. These factors include the availability of a wide range of nutrients in the mouth, presence of different types of habitats and also the formation of plaque on teeth; thus providing an environment to bacterial communities to survive in close proximity to each other. Since 1960, various studies on bacterial diversity have shown the changes in the formation of plaque. Ritz (1967) studied the composition of microbes in supragingival plaque over the period of nine days and found that on day one *Streptococcus* and *Neisseria* were the more prevalent forms of microbes. After nine days the proportion of aerobic bacteria decreased and there was an increase in the number of anaerobic bacteria e.g *Veillonella, Corynebacterium* and *Fusobacterium* (Ritz, 1967). During the initial phase of biofilm formation, Gram-positive cocci were mostly observed and overnight a thick layer of biofilm was observed containing different forms of coccoid, rod-shaped, filamentous and coco-bacillary bacteria. *Streptococcus* and *Actinomyces* were also found to be the most prevalent forms among the early colonizers (Kolenbrander *et al.*, 2006; Li *et al.*, 2004; Nyvad & Kilian, 1987).

1.4.1 Bacterial Diversity Observed Using Cultivation Methods/Phenotypic Analysis

The diversity and composition of the microbial flora has been investigated using different culture-dependent and molecular approaches. The process of scientific naming to bacterial isolates has long been used in the field of clinical microbiology. Initially microbiologists compared the bacterial isolates based on either their morphological and phenotypic descriptions found in the *Bergey's Manual of Systematic Bacteriology* or the *Manual of Clinical Microbiology* (Lehmann, 1999; Scannapieco *et al.*, 1989), or compared their unknown isolates based on the published tables by the well renowned *Centre for the Disease Control* and *American Type Culture Collection* (ATCC) centers,

where these well-characterized bacterial species were kept. Only 50 % of oral micro-organisms have been cultivated so far (Haffajee *et al.,* 2008; Wilson, 2005). There were about 280 bacterial species in the oral cavity which have been identified by culture methods (Dewhirst *et al.*, 2010). The conventional culture methods to cultivate the micro-organisms were the essential methods to observe the bacteria in the past. With the advent of anaerobic cabinets and anaerobic jars, there were advances in the field of isolation of fastidious anaerobic isolates, although anaerobes are slow growers and require specialized media and incubation temperatures, however using these techniques 250 more micro-organisms were successfully isolated. The specificity for nutritional and physico-chemical requirements for each micro-organism is the main reason which poses difficulty in growing the micro-organisms in the laboratory environment.

The oral bacterial population is unique in each individual. The most prominent cultivable microflora (Table 1.1) isolated from the healthy human palatal mucosal surface were *Streptococcus* (52%), *Actinomyces* (15%), *Lactobacillus* (1%), *Neisseria* (2%), *Veillonella* (1%), *Prevotella* (4%) and also *Candida* (isolated in very low numbers) (Gusberti *et al.*, 1985; Marsh *et al.*, 2009). The commonly isolated micro-organisms from buccal mucosa using cultivated techniques in the humans were *S. sanguinis, S. salivarius, S. oralis, S. mitis* along with few mutans streptococci, *Actinomyces naeslundii, A. odontolyticus, Capnocytophaga* spp., Fusobacterium spp. and black pigmented colonies and *Haemophilus* spp. The dorsum of the tongue was found to be the most diverse habitat of the mouth and it harbored a high frequency of the micro-organisms due to its large surface area and papillated surface. Streptococci (51.7%) were cultivated most numerously and especially the *salivarius/mitis* group was in abundance along with *S. anginosus, S. oralis, S. mutans* and *S. sobrinus*. The *Neisseria* (20%), *Veillonella* spp. (6.3%), *Rothia mucilaginosa* (5.5%), *R. dentocariosa* (0.9%), *Actinomyces* (5%), and also species of Gram-negative bacteria like *Haemophilus*, *Fusobacterium*, *Prevotella*, *Capnocytophaga* and *Aggregatibacter* have also been recovered using cultivation techniques (Marsh *et al.*, 2009; Marsh & Percival, 2006). The teeth provide the hard surface for the attachment of bacteria ultimately forming plaque or biofilms in the oral cavity. Dental plaque was also known by qualifying terms based on the site from where

it was recovered such as smooth surface, approximal, fissure, gingival crevice, supragingival and subgingival plaque. *Streptococcus* and *Actinomyces* are the major organisms cultured from plaque. Obligate anaerobes and spirochaetes were also found to be present in the plaque sample (Li *et al.*, 2004; Marsh *et al.*, 2009; Roberts, 2005).

The list of major Phyla, respective genera and their species found in the oral cavity

Phyla, Genera	**Species**
Firmicutes	
Anaerococcus	*A. prevotii*
Catonella	*C. morbi*
Centipeda	*C. periodontii*
Dialister	*D. invisus, D. pneumosintes*
Eggerthella	*E. lenta*
Enterococcus	*E. faecalis*
Eubacterium	*E. sulci, E. infirmum, E. saphenum, E. nodatum, E. brachy, E. minutum*
Filifactor	*F. alocis*
Finegoldia	*F. magna*
Gemella	*G. morbillorum*
Granulicatella	*G. adiacens*
Lactobacillus	*L. salivarius, L. acidophilus, L. fermentum, L. paracasei, L. catenaformis*
Mogibacterium	*M. timidum, M. pumilum, M. neglectum, M. vescum*
Parvimonas	*P. micra*
Peptoniphilus	*P. asaccharolyticus, P. lacrimalis*
Peptostreptococcus	*P. anaerobius*
Pseudoramibacter	*P. alactolyticus*
Selenomonas	*S. sputigena, S. noxia*
Solobacterium	*S. moorei*
Streptococcus	*S. mutans, S. sobrinus, S. mitis, S. sanguinis, S. gordonii, S. oralis, S. anginosus, S. constellatus, S. intermedius, S. salivarius*
Veillonella	*V. parvula*
Bacteroidetes	
Capnocytophaga	*C. gingivalis, C. ochracea*
Porphyromonas	*P. endodontalis, P. gingivalis*
Prevotella	*P. intermedia, P. nigrescens, P. tannerae, P.multissacharivorax, P. baroniae, P. denticola,*
Tannerella	*T. forsythia*

Actinobacteria

Actinomyces	*A. israelii, A. gerencseriae, A. naeslundii, A. oris, A. johnsonii, A. meyeri, A. odontolyticus*
Atopobium	*A. parvulum, A. minutum, A. rimae*
Bifidobacterium	*B. dentium, B. adolescentis, B. bifidum*
Corynebacterium	*C. matruchotii*
Olsenella	*O. uli, O. profusa*
Propionibacterium	*P. acnes, P. propionicum*
Rothia	*R. dentocariosa*
Slackia	*S. exigua*

Proteobacteria

Aggregatibacter	*A. actinomycetemcomitans, A. aphrophilus*
Campylobacter	*C. rectus, C. gracilis, C. curvus, C. showae, C. concisus*
Eikenella	*E. corrodens*
Neisseria	*N. mucosa, N. sicca*

Fusobacteria

Fusobacterium	*F. nucleatum, F. periodonticum*
Leptotrichia	*L. buccalis*

Spirochaetes

Treponema	*T. denticola, T. socranskii, T. parvum, T. maltophilum, T. lecithinolyticum*

Table 1.1: The list of major Phyla, respective genera and their species found in oral cavity

(Taken from (Kielty, 2011) using cultivation methods)

1.4.1.1 Phenotypic Differences Between Cultivated Species

The oral species which were isolated using culture methods (Table 1.1) may be differentiated using phenotypic methods e.g colony morphology on solid media (colour, size, shape, opacity, elevation, margin, surface texture, consistency etc), Gram staining and biochemical identification etc. The classification of oral micro-organisms based on phenotypic differences is very tedious and requires more comprehensive information regarding specificity of nutrient uptake of micro-organism in media etc. There were many bacteria involved in plaque formation but since the 1990s, the emphasis was given to two main genera being as pioneer bacterial genera.

1.4.1.1.1 *Streptococcus*

Streptococci constitute the largest group of cultured micro-organisms from all sites of the oral cavity. Streptococci show partial haemolysis on blood agar, and early researchers based on phenotypic features classified them into a group of "viridans streptococci" however more recently they are clustered into four groups (Beighton *et al.*, 1991b; Whiley, 1987; Whiley *et al.*, 1990). The mutans group contains species called *S. mutans*, *S. sobrinus, S. cricetus, S. rattus, S. downei, S. macacae* and *S. ferus*. This group is very important because of its role in the etiology of dental caries (Marsh *et al.*, 2009; Whiley & Beighton, 1998). In 1924, *S. mutans* was isolated by Clarke from carious lesions in human teeth and then isolated from a patient of infective endocarditis. The mutans streptococci showed differential colony morphologies on plates containing sucrose due to their ability to produce soluble and insoluble extracellular polysaccharides (glucan, mutan and fructan) from sucrose and finally converts them into acids like lactates (Shah & Russell, 2004; Wiater *et al.*, 2012). Salivarius-group contains *S. salivarius* and *S. vestibularis.* They are present mostly on the tongue surface. *S. salivarius* grew as mucoid colonies on sugar-containing agar plates due to its characteristic feature of levan production from starch. *S. vestibularis* produce urea which has the ability to increase the pH. The Anginosus-group contained species named *S. constellatus, S. intermedius, S.*

anginosus and they all were found to be present in dental plaque and on mucosal surfaces. They cause significant maxillo-facial infections in humans. They were also found in abscesses of internal organs, endocarditis and meningitis. They produce toxins that can affect the host defense system. The mitis-group contains many species. *S. gordonii* can break down starch due to its ability to bond to alpha amylase of starch. *S. oralis* produce neuraminidase and IgA protease (Kuvatanasuchati *et al.*, 2012; Liebana *et al.*, 1993; Marsh *et al.*, 2009).

1.4.1.1.2 *Actinomyces*

Actinomyces are the second largest group of micro-organisms found in dental plaque at interproximal sites and in gingival crevices. *Actinomyces* are short rods and pleomorphic in shape. A few species show a branching morphology when observed under the microscope. *Actinomyces* have the ability to metabolize glucose into succinic, acetic and lactic acids and this feature was used in the initial identification of *Actinomyces* species (Takahashi & Yamada, 1999). *A. israelii* are filamentous, *A. naeslundii* are fimbriated while the rest of the species have a smooth surface. The flow chart (Figure 1.11) was drawn for the initial identification of *Actinomyces* (Sarkonen *et al.*, 2001).

A. oris (previously known as *A. naeslundii*) was the most common in dental plaque. *A. naeslundii* was classified into two genospecies. Some strains of this group produce enzymes that can hydrolyze fructan into levans and inulin and some have the ability to produce urease which is an important enzyme to change the pH in plaque (Marsh *et al.*, 2009). Fimbriae are the special adhesion structures found in this group of *Actinomyces* and they play a role in cell-to-cell contact and cell-to-surface contact when colonizing a dental surface (Mishra *et al.*, 2010). Fifty percent of strains of *A. odontolyticus* form red-brown pigmented colonies (Sarkonen *et al.*, 2000) but phenotypic tests are of little use in identifying *Actinomyces* to the species level.

.

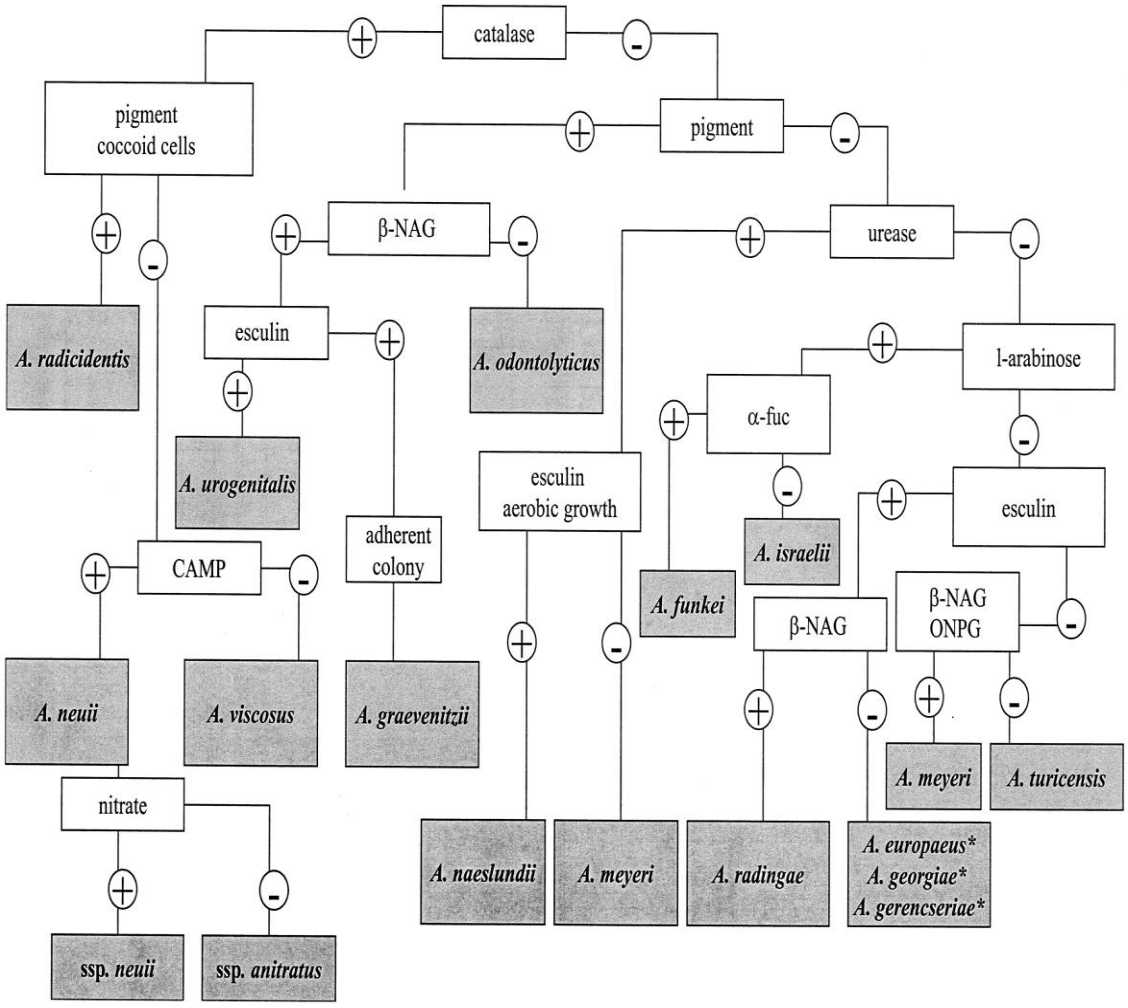


Figure 1.11: A flow chart for initial identification of *Actinomyces* species (taken from Sarkonen *et al.*, (2001)).

β-NAG, β-N-acetyl-glucosaminidase; ONPG, o- nitrophenyl-β-D-galactopyranoside; CAMP test, synergistic hemolysis; α-fuc, α-fucosidase.

1.4.2 Genetic Diversity In Bacteria

Point mutation, horizontal gene transfer and recombination are the basic mechanisms by which a change in a bacterial genome is possible and which result in increased genetic diversity.

1.4.2.1 Point Mutation

A point mutation or single base mutation was considered responsible for evolutionary change in a population. Bacterial genome was changed by acquisition or loss of genetic sequence. During DNA replication, the substitution of a wrong base was called mutation. Point mutations were either transition or transverse type depending on the nature of base change. Transitions referred to change a purine base with another purine base and vice versa, while the transverse type is opposite as the purine base was changed in place of pyrimidine base and vice versa. The change in nucleotide base depending on functional role was of two types: Synonymous substitutions (silent mutation) in which no change in amino acid residue occurred therefore the protein sequence remained the same, while in non-synonymous mutations a change in protein sequence occurred which finally results into a truncated gene product which were functionally impaired (Hardie, 1992).

1.4.2.2 Horizontal Gene Transfer

Horizontal or lateral gene transfer terms were used when bacteria acquired the genetic material from other than the immediate ancestor. This is the most widely known method in nature to introduce diversity among bacterial population (Gogarten & Townsend, 2005; Nakamura *et al.*, 2004; Thomas & Nielsen, 2005). A study, which involved the nucleotide composition analysis showed that among 116 prokaryotic genomes, 14 % of the genes, had undergone horizontal gene transfers (Nakamura *et al.*, 2004). A study on *E. coli* represented that 40%-50% of the genes were common among all selected strains, while the rest of the genes evolved from potential horizontal gene transfer and resulted

into the formation of pathogenicity islands arrangements (Lloyd *et al.*, 2007; Welch *et al.*, 2002). Pathogenicity islands were explained as 'horizontally transferred genes which contributed in virulence of bacteria'.

1.4.2.3 Recombination

Recombination is defined as the exchange of genetic information between two similar DNA molecules. The homologous recombination occurred in closely related bacterial species but very rarely (Selander & Levin, 1980). Homologous recombination was accepted widely after its influence on population structure was shown in a study on *E. coli* (Feil & Spratt, 2001; Wirth *et al.*, 2006). The studies showed that change in a housekeeping gene was more likely to occur due to recombination rather than mutation in bacterial species like *Neisseria meningitidis*, *Streptococcus pneumoniae*, and *Staphylococcus aureus* (Feil *et al.*, 2000; Feil & Spratt, 2001; Spratt & Maiden, 1999). Association between genes at different loci can be clonal and panmictic. There are isolates of bacterial species which are not easily distinguishable in genotypes are known to have clonal mode of recomibation. The extent of recombination among clonal complexes varies among bacteria. Some species have stable clones (e.g. *Salmonella enterica*), whereas other species (e.g. *Helicobacter pylori*) have transient clones. Panmictic recombination is opposite to clonal. In panmictic mode of recombination the genetic diversity is very high due to high levels of recombination (e.g *N. gonnhorrea* and *P. falciparum*) and sexual recombination is most frequent (Anderson *et al.*, 2000; Smith *et al.*, 1993).

1.4.3 Molecular Analysis Of Oral Microbial Diversity

The failure of conventional methods for culturing micro-organisms was the basis to use molecular methods to identify micro-organisms which were not-yet-culturable (Vartoukian *et al.*, 2010). It has been estimated that approximately half of the total human oral microbiome is as-yet-uncultivated (Dewhirst *et al.*, 2010). Oral microbial diversity was investigated in detail in the past using culture-independent/molecular methods. The innovative idea of using the molecular techniques was illustrated by Zuckerkandl and Pauling (1965). According to authors, the evolutionary relationship among micro-organisms can be inferred by collecting the information on biological macromolecules (Zuckerkandl & Pauling, 1965). This concept emerged into a novel field of molecular phylogeny. The second main reason of using a molecular approach is to present a precise nomenclature of isolates based on exact phylogenetic position. The research community was continuously isolating new species and was using the National Center for Biotechnology Information (NCBI) search engine to investigate the bacterial phylogeny but an NCBI search was not always very helpful due to lack of good reference sequences for the majority of genera (*Actinomyces, Bifidobacteria etc*) on the public database. Therefore, molecular methods, particularly culture-independent methods, have increasingly been used in studies of the human oral microbiome. The well established molecular techniques of PCR, 16S rRNA gene sequencing (Choi *et al.*, 1996; Clarridge, 2004), and cloning were used to identify bacterial species for taxonomic studies (Dewhirst *et al.*, 2010). 16S ribosomal RNA genes were used most commonly (Woese, 1987). The RNA/DNA extracted from whole micro-organisms was sequenced and aligned using mathematical models with phylogenetic trees being constructed to show the evolutionary relationship among the micro-organisms. The house keeping genes were also used as a key technique to observe the diversity. The idea behind the use of housekeeping genes was that these genes were the conserved regions on DNA and can be used easily for the alignment of sequences of related micro-organisms. The representative phyla which were found in the oral cavity includes *Actinobacteria*, *Firmicutes*, *Fusobacteria*, *Spirochaetes*, *Proteobacteria*, TM7 and *Bacteroidetes* using molecular

based techniques (Dewhirst *et al.*, 2010). The cultivation-independent methods like 16S rRNA gene-based cloning studies have also been used to identify almost 700 common bacterial species (Dewhirst *et al.*, 2010).

1.4.3.1 Oral Diversity Using 16s rDNA Analysis

The community of micro-organisms residing in the oral cavity varies between individuals and is unique to each individual (Aas *et al.*, 2005; Bik *et al.*, 2010; Mager *et al.*, 2003). Oral bacteria have been a major focus of interest when medically important communities were analyzed using 16S rRNA-based techniques, cloning and sequencing (Beighton, 2005; Choi *et al.*, 1994; Dewhirst *et al.*, 2010; Dymock *et al.*, 1996; Flores *et al.*, 2012; Harper-Owen *et al.*, 1999; Hugenholtz *et al.*, 1998; Jung *et al.*, 2000; Kroes *et al.*, 1999; Nyvad *et al.*, 2013; Paster *et al.*, 2001; Rolph *et al.*, 2001; Sakamoto *et al.*, 2000; Spratt *et al.*, 1999; Tanner *et al.*, 1994; Wade *et al.*, 1997; Ximénez-Fyvie *et al.*, 2000). The 16S rRNA gene was believed to be ideal for taxonomic assignment of organisms because of the presence of a combination of conserved and hypervariable regions of sequence (Clarridge, 2004). The amplification of 16S rRNA genes was carried out using a well-implemented methodology and for this purpose 'universal' or broad range primers were used. The resulting mixture of gene sequences were cloned and separated using competent *Escherichia coli* cells and subsequently the cloned inserts were sequenced. The sequences can then be compared to the available reference database for identification such as Human Oral Microbiome Database (HOMD) (Chen *et al.*, 2010), The ribosomal Database project (Cole *et al.*, 2009) or the NCBI nucleotide database using BLAST (Altschul *et al.*, 1990). This methodology can identify a wide range of species present in a sample, including the 'unculturable' representatives of the oral microbiome. The first large study was undertaken to resolve the issue of naming and isolation of oral micro-organisms by Paster *et al.*, in (2001) and investigated the bacterial diversity in the oral cavity by using culture independent methods for uncultivated or not-yet-cultivated oral micro-organisms. There were about 2,552 clones analyzed and complete sequences

containing 1500 bases were obtained for new species. There were about 500 species or phylotypes which were investigated and named during this study (Paster *et al*., 2001).

In a study by Dymock *et al*.,(1996), five groups of micro-organisms were identified, which were un-cultivated previously. The technique, which they used, was direct amplification of 16S rRNA or DNA genes, cloning and finally sequencing. The species found from dentoalveolar abscesses were *Porphyromonas gingivalis, Prevotella oris, Peptostreptococcus micros, Zoogloea ramigera,* and one was distant to *Prevotella*. Munson *et al* (2002) did a study on endodontic infections and they used both cultural and molecular techniques to identify the micro-organisms having investigated a total of 261 isolates and 624 clones. 20 taxa were found per sample, which was more than recorded before with only culture methods while the *Dialister invisus* was the only species commonly found in all samples. The predominant phyla were *Firmicutes, Bacteroidetes,* and *Actinobacteria.* A study based on finding diversity among *Spirochaetes* showed that most of the species of this group were uncultivated. They used the 16S rRNA gene sequencing technique with universal primers, cloned and constructed many libraries. A *Treponeme* specific probe was used. They found about 23 new species from a single patient (Choi *et al.*, 1994). *Streptococcus* species was subsequently classified into five groups using 16S rRNA sequences by Kawamura (1995) and a Neighbor-Joining tree was produced (Figure 1.12). Aas *et al*., (2005) undertook a similar study to find the diversity of oral *Spirochaetes* in patients with peridontitis. Dewhirst (2010) used 16S rRNA gene sequencing technique using selective primers for identifying *Spirochaetes* on plaque samples extracted from periodontal pockets of humans. 10 known and 47 as-yet-uncultivated *Treponema* species were isolated thus providing the prevalence of this genus in oral diseases. In a study by Clarridge (2004), a universal phylogenetic tree was made using the 16S rRNA sequencing technique as shown in Figure 1.13.

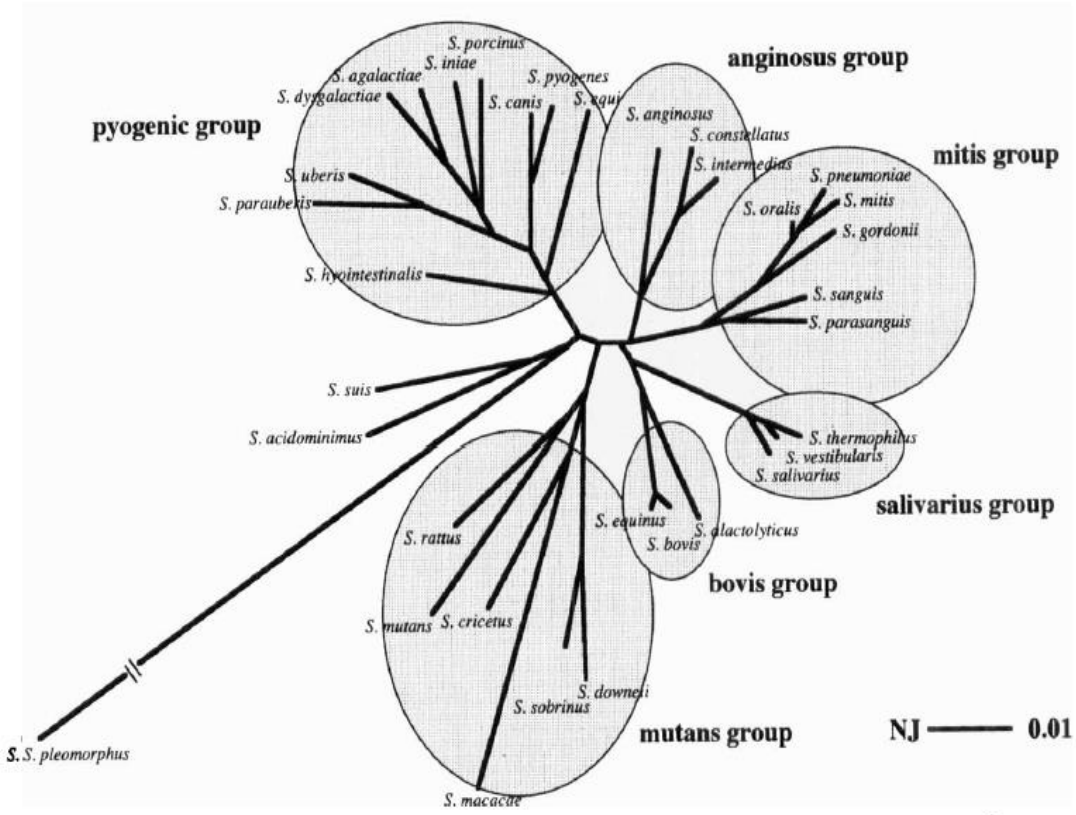


Figure 1.12: Comparison of 16S rRNA sequences of individual cultivable streptococcal species (Kawamura *et al*., 1995)

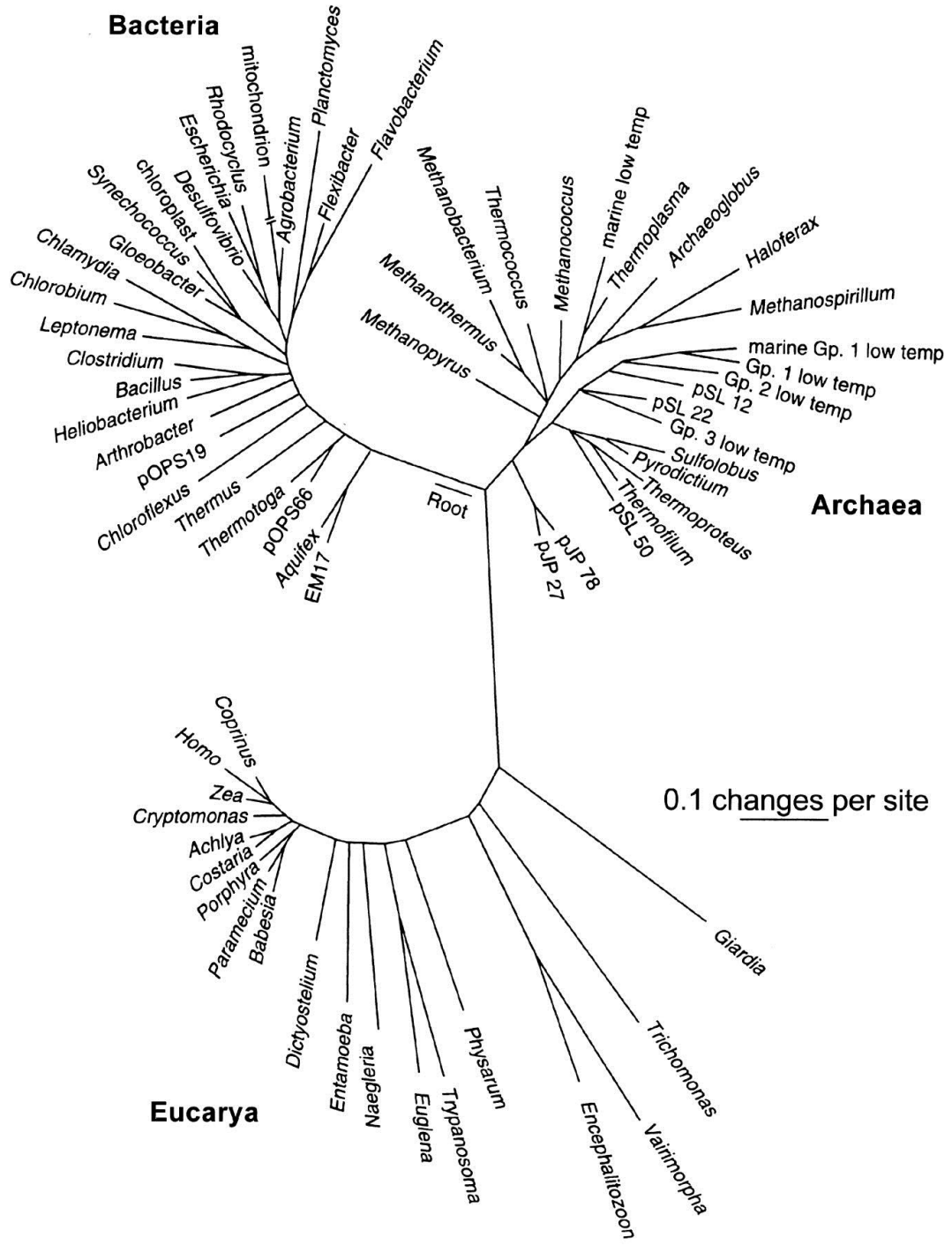


Figure 1.13: Universal Phylogenetic tree (16S rRNA gene sequence comparisons) (Clarridge, 2004)

1.4.3.1.1 Limitations Of 16S Analysis Methods

The 16S rRNA gene PCR, cloning and sequencing studies were not always ideal due to the limitation of the methodology adapted in implementing the technique. The biases came up due to the use of 'universal' PCR primers, variability in numbers of rRNA operons between species, cloning and DNA extraction methods (Nossa *et al.*, 2010). The DNA extraction methods for Gram-positive bacteria are not always adequate, as they require special lysis treatment to break the cell wall of bacteria. PCR and the number of cycles in the PCR sometimes introduced biases in the results (Bonnet *et al.*, 2002). The high G+C content phylum of *Actinobacteria* was also under-represented in culture-independent molecular cloning studies (de Lillo *et al.*, 2006). The high G+C organisms [Bifidobacteria and Actinomyces] were missed during 16S rRNA sequencing (Dewhirst 2010). The numbers of 16S operons differ between genera/species, and not all 16S operons in the same strain are necessarily identical. The full 16S sequences did not enable differentiation of species of some genera and often sequencing data was insufficient to identify species.

1200 taxa were defined at the species/phylotype level using 16S rRNA sequencing (Aas *et al.*, 2005). The diversity was minimum using 16S rRNA sequencing studies. The question was raised that were all members of the same species or phylotypes identical. Different molecular techniques other than 16S rRNA sequencing were used to differentiate the isolates at species levels and the techniques mostly involved were phenotypic analysis, DNA fingerprinting, comparison of individual gene sequences, Multi-locus sequence typing [MLST], and whole genome comparisons etc.

1.4.3.2 DNA Fingerprinting [REP-PCR]

DNA fingerprinting was another popular method used to observe the diversity at species level. A number of studies were done using this technique to delineate the micro-organisms at genus or species levels. Repetitive extragenic palindromic PCR [REP-PCR] is a molecular method, which is very useful for quick grouping and identification of bacteria. The DNA of micro-organisms has repetitively inverted DNA elements present on the entire genome. The primers were designed for these repetitive sequences and PCR was done to amplify and run them on gel electrophoresis. Same species could be clustered together using this rapid method of identification. In one study, the PCR-based method of DNA fingerprinting was applied successfully to distinguish within the viridians group of streptococci (Alam *et al.*, 1999). In another study using 16S rDNA and then recA housekeeping gene sequence comparison revealed the presence of three types of *P. acnes* phylogenetically (McDowell *et al.*, 2008). The use of REP-PCR based technique in a study by Niazi *et al.*, (2010) demonstrated that *P. acne* isolated from endodontic infections were significantly different from the isolates of skin. The recA house keeping gene sequence analysis further revealed the presence of three types among *P. acne* isolates, type I was originally from skin and type II and III were from endodontic infections. Genotypic heterogeneity was observed among the strains of *Streptococcus oralis* using REP-PCR technique (Alam *et al.*, 2000). This was the first report of genotypic heterogeneity amongst non-mutans streptococci relating to aciduricity in dental plaque. The REP-PCR-based method was also applied successfully to investigate the effect of environment on genotypic diversity of *Actinomyces naeslundii* and *Streptococcus oralis* in the oral biofilm (Paddick *et al.*, 2003). In a recent study by Cheon *et al.*, (2011), the genetic diversity of plaque mutans streptococci was dissected using the REP-PCR-based technique.

1.4.3.3 MLST Analysis

The Multi Locus Sequence Typing (MLST) has also been used to study the oral diversity at species level. The advantage of this type of molecular analysis is that, it has sequence data which is compatible with a portable typing system. Moreover, it permits strain comparison between remote laboratories. It is easily accessible on-line and the data subject to phylogenetic analysis is useful for epidemiological studies [acquisition, cross-infection, pathogenic and epidemic strain identification]. For each allele, different allelic sequences are designated a number and each different combination of alleles is designated a sequence type [ST]. MLST (Multi locus Sequence Typing) is a molecular technique implemented in 1998 (Maiden *et al.*, 1998) due to the limitations of the already present 16S rRNA and other related techniques. The MLST technique was successfully used by researchers to differentiate the micro-organisms at a high resolution at sub-species level for use in molecular-epidemiological studies. The partial sequences of multiple 'housekeeping' genes were used in the MLST. MLST has, however, also been proposed as a tool for reliable differentiation of bacterial species (Hanage *et al.*, 2005; Segata *et al.*, 2012), particularly in cases where 16S rRNA gene sequencing alone, is insufficient. Gevers *et al.*, (2005), proposed that the term Multi-locus Sequence analysis should be used in place of MLST when applying the technique to the identification of bacteria at the species level. The approach used initially was 16S rRNA to group the bacteria to the genus or family level and subsequently primers were designed for MLS analysis (MLSA) to allow the identification to the species level. Hanage (2005) concluded that individual loci were inadequate for differentiating species of closely related species, whereas multi locus approach could be considered more effective. This was demonstrated using phylogenetic trees of concatenated sequences of 'housekeeping' genes used in the MLST scheme, that were amplified from species of bacteria.

Bifidobacteria was considered to be a prominent member in the digestive tract but little was known about their presence in the oral cavity (Beighton *et al.*, 2008). However the bifidobacteria are commonly misidentified to actinomyces because of being catalase negative and Gram-positive rods, therefore the *Bifidobacterium* was always confused by

taxonomists with *Actinomyces* when identifying using culture methods. Therefore it was necessary to use rRNA sequencing methods to distinguish *Bifidobacterium* from *Actinomyces* (Mahlen & Clarridge, 2009). The *Actinomyces* were further delineating successfully to sub-species level using the MLST approach in a study in this laboratory by Henssge *et al*., (2009).

Henssge *et al*., (2009) demonstrated the use of MLST and based on their results an emended description of *Actinomyces naeslundii*, *Actinomyces oris* and *Actinomyces johnsonii* was produced. These species were previously known as genospecies 1, 2 and WVA 963 respectively.

Actinomyces naeslundii was found to be present as an early colonizer in the oral biofilm. It is not easy to designate the exact phylogenetic position to the isolates belonging to this species using conventional phenotypic tests and it is not even very well differentiable using 16S rRNA sequencing techniques. The MLST approach was adopted for the 2009 study to differentiate the genospecies of *Actinomyces* and for this purpose seven house keeping genes were used for MLST analysis. *Actinomyces naeslundii* was found to fall into three genospecies using this molecular-identification approach (Henssge *et al.*, 2009). Thus the designation of species belonging to this genus was emended using MLST technique.

In a study by Do *et al*., (2010) on *Streptococcus mutans,* genetic diversity was observed by MLST and found that this species had a very low genetic diversity and Sequence Types (ST's) are readily distinguished by MLST. The Phylogenetic tree was produced based on MLST analysis using 6 house keeping genes [*acc*C (acetyl-CoA carboxylase biotin carboxylase subunit), *gki* (glucokinase), *lepA* (GTP-binding protein), *recP* (transketolase), *sodA* (superoxide dismutase), and *tyrS* (tyrosyl-tRNA synthetase)] and ClonalFrame program was used to observe the recombination/mutation rate (Figure 1.14).

In conclusion, all isolates of the same species are different among members of the normal commensal flora. The microbial diversity in the oral biofilm describes minimum diversity using 16S rRNA sequencing, and DNA fingerprinting increases the diversity and MLST approach greatly increases diversity.

The microbiological research concepts were initially based on investigating one organism and observation obtained called the "reductionism" theory. Recently this phenomenon has been changed significantly and researchers started focusing on understanding the whole also known by the term "holism". Microbiologists are currently using approaches including global gene regulation and metagenomics to understand the interactions of micro-organisms when found in a community or biofilm (Kuramitsu *et al.*, 2007).

The oral micro-flora contains a diverse community of micro-organisms which together forms a microbial complex or biofilms. This diverse community of micro-organisms interact with each other and with the host cells forming a 'microbial ecosystem'. There is much ongoing research to isolate and identify these microbial species and now the research has moved a step forward to attempt to understand the role of certain micro-organisms. The answer to these questions could be elucidated by a greater understanding of the metabolic activities possessed by oral micro-organisms and this could give a clue to control the microbial ecosystem in terms of controlling oral diseases such as dental caries, periodontal diseases and oral malodour (Takahashi, 2005).

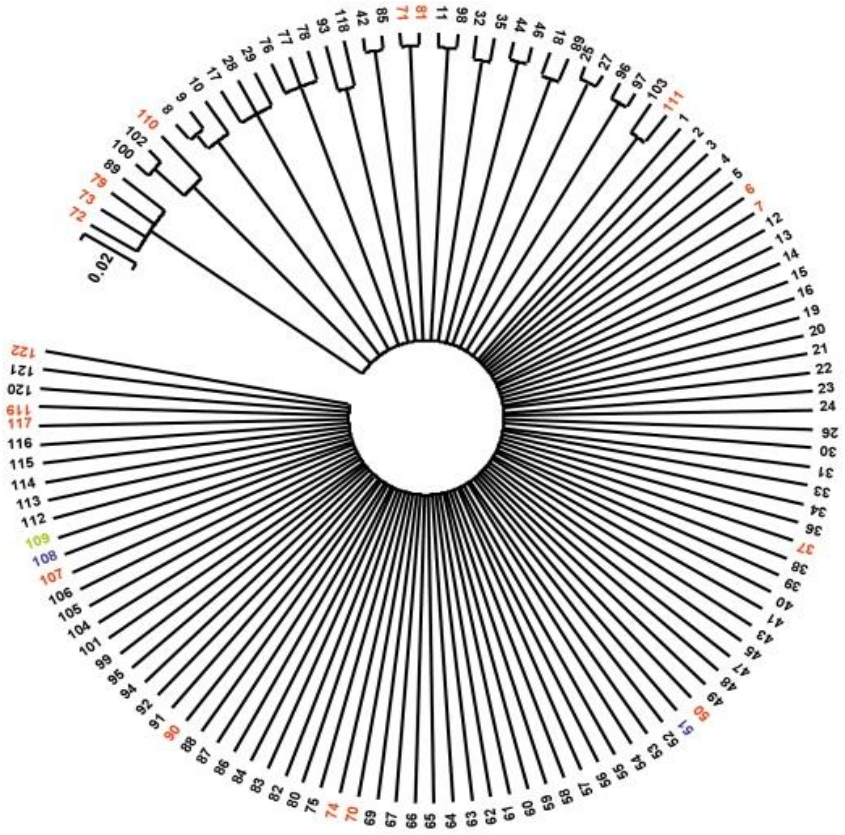


Figure 1.14: Radiant Phylogenetic tree of *S. mutans* (122 STs) (Do *et al.*, 2010)

***The clonal relationship of *S. mutans* population at 6 loci was shown. Serotypes c, e, f and k represented with black, red, blue and green colour numbers respectively.**

1.5 Genome Sequencing

The Whole genome sequencing is the most advanced technique introduced to examine the phylogenetic association of bacteria. The advent of modern state-of-the art high-throughput sequencing technologies facilitates the tedious processes of DNA sequencing and allowed researchers to study the million base pair read data in a single experimental run and in a cost effective way (Brenner *et al.*, 2000; Mitra & Church, 1999).

1.5.1 Roche 454 (Pyrosequencing)

Roche 454 is the first commercially available Next Generation Sequencing (NGS) platform. It is also known as Genome Sequencer GS (FLX). It works on the principle of pyrosequencing. This principle was first described by Ronaghi (1996). The system instead of using fluorescently labeled nucleotides or primers, worked by converting pyrophosphate to ATP that, in turn converted to light by using firefly luciferase and this can be coded into readable sequences. The Roche 454 platform was first introduced in 2005 after long optimization (Margulies *et al.*, 2005). The GS (FLX) platform has the ability to give 5 Mbases in a single run of 5h using titanium reagents. A bacterial genome of 20 million base pairs length can be *de novo* sequenced with 25x coverage and data was enough to assemble correctly 95-99% of sequences. The read length obtained with *pyrosequencing* is the main advantage of the GS (FLX) system. The reads obtained with the GS (FLX) system can be assembled into a few hundred contigs. The commercially available software is easily used to assemble the draft genome consisting of a few to a few hundred contigs within a few hours. This data was sufficient for many genomic and taxonomic studies.

1.5.2 Illumina Genome Analyzer

The Illumina Genome Analyzer (IGA) is a platform that has been used to obtain millions of very short sequences of 36, 76 or 100 bases in length based on the model of the system used. These short reads are assembled into larger contigs using proprietary software into exploitable data. The denovo assembly programs generated by different software solution companies are available online to perform such tasks (Hernandez *et al.*, 2008). The principle of Illumina Technology is based on sequencing-by-synhesis. Genomic DNA was fragmented and adapters were ligated at both ends of the DNA fragments. The DNA fragments were then attached to inside surface of the flow channel. Unlabelled nucleotides and enzymes were added at this stage to initiate the formation of bridge which will in turn incorporate double stranded nucleotides and several million clonal clusters of DNA fragments was formed as shown in Figure 1.15. The first sequencing cycle was initiated by adding all four labelled reversible terminators, primers and DNA polymerase enzymes. The laser excitation was used to capture the image of emitted fluorescence from each cluster on the flow cell. The identity of first base was recorded in each cluster. Similarly the next sequencing cycle was initiated and image of the second base was recorded in a similar way. Sequence of bases was determined by repeating the cycles of synthesis in a given fragment. Images captured per each nucleotide attached gives different coloured fluorescence and from this the sequence of necleotide addition can be determined (Figure 1.16).

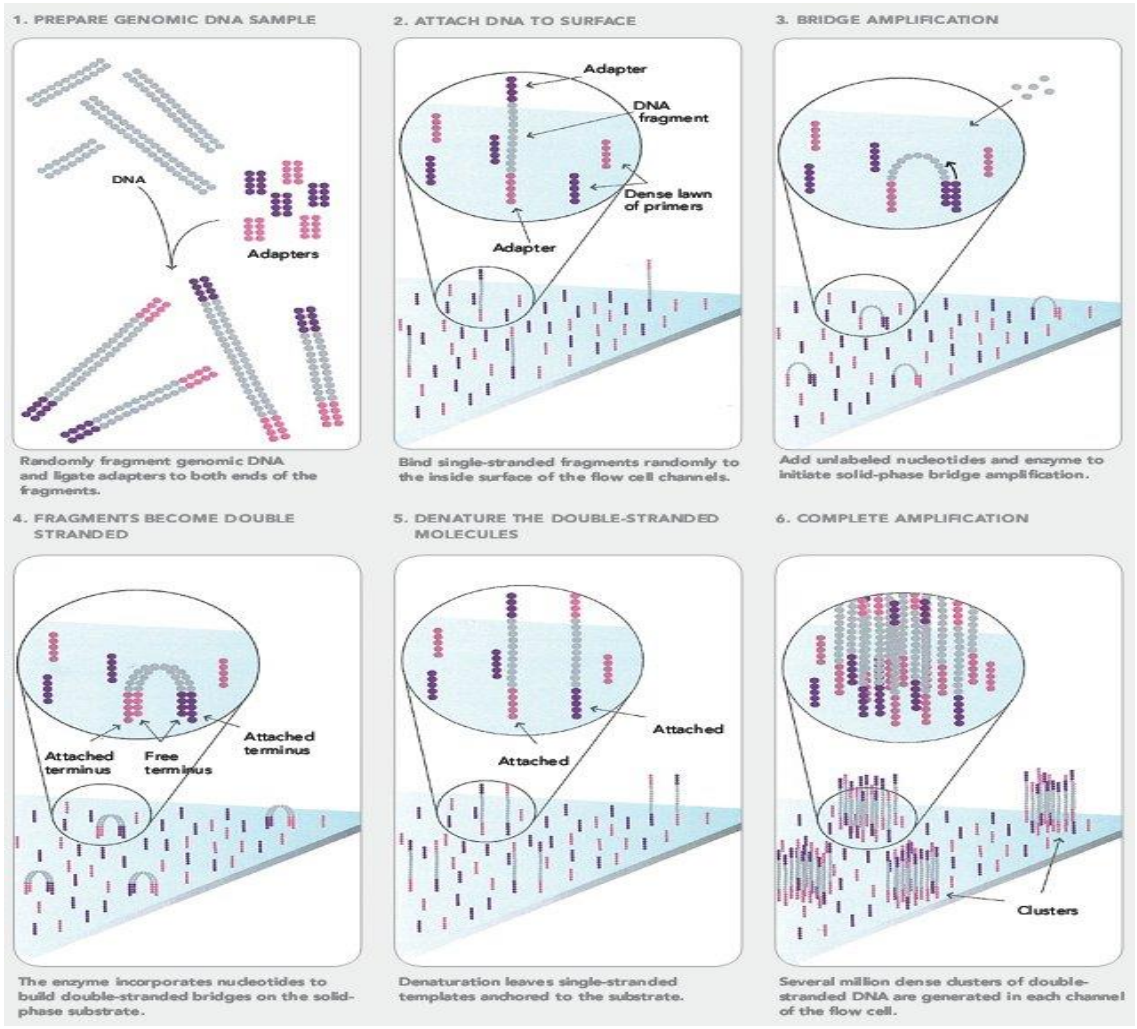


Figure 1.15: DNA cluster formation by bridge amplification on the flow cell using Illumina Genome Analyzer sequencing technology.

***(Taken from website http://res.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf)**

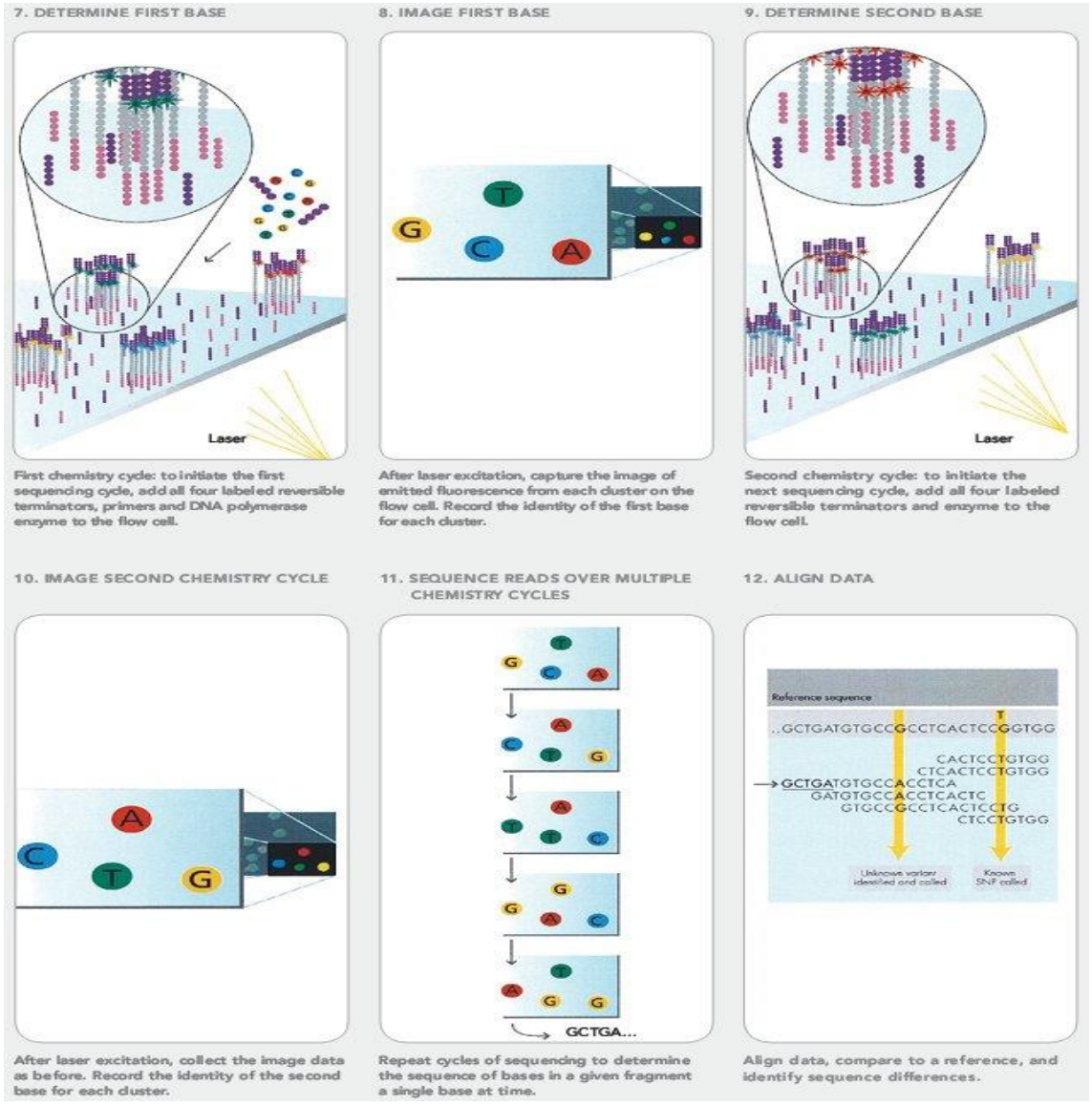


Figure 1.16: Sequencing-by-synthesis principle steps using Illumina Genome Analyzer sequencing technology.

***(Taken from website http://res.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf)**

1.6 *Actinomyces*

Actinobacteria is one phylum of the oral microbiome and *Actinomyces* is a genus within *Actinobacteria.* These are Gram positive, anaerobic or facultative anaerobic, rod-shaped bacteria and have a high G+C content. *Actinomyces* are considered as normal microflora of the mouth but have also been investigated as etiologic agents of certain diseases such as classical actinomycosis, eye infections, abscesses at different body sites, oral, genital and urinary tract infections (Sarkonen *et al.*, 2001; Schaal & Lee, 1992).

The involvement of *Actinomyces* in the caries process was studied and has shown that *Actinomyces* were heterogeneous in respect of acidogenicity (Nyvad & Kilian, 1987). Dental caries is destruction of the tooth by demineralizing the hard tissues to form a carious lesion (Fejerskov & Kidd, 2008). *Actinomyces* were found to grow at low pH and 21.6% of total microbial flora of root surface caries lesions grew at pH 4.8 and 10.7 % amongst them were *Actinomyces* (Brailsford *et al.*, 2001). A study reported that isolates collected from different sites in the oral cavity have significantly higher number of *A. naeslundii* (genospecies 2) currently known as *A. oris* as compared to *A. naeslundii* (genospecies 1) (Brailsford *et al.*, 1998). Another study reported that the 92% of isolates collected from patients suffering from infected dentine of active root caries lesions were identified as *Actinomyces*, 17.9% of isolates were identified as *A. naeslundii* (genospecies 1) and others were *A. naeslundii* (genospecies 2) (Brailsford *et al.*, 1999). These taxa were recently termed as *A. naeslundii* and *A. oris*, respectively by Henssge *et al*., (2009).

1.6.1 *Actinomyces* And Closely Related Genera

The exact phylogenetic position of *Actinomyces* has received much attention within the scientific community. In 1877, *Actinomyces bovis* was the first micro-organism discovered when the German veterinarian named Otto Bollinger investigated the chronic destructive disease in cattle, which involves tongue and jaw and it was previously confused with sarcomatous malignant tumour (Bollinger, 1877). *A. bovis* was observed

using a microscope and appeared to contain small crystal like particles called “drunsen” or “vugus” later known as sulphur granules. The unique pattern of growth, like ray-fungus, lead the scientist to give this organism the name *Actinomyces bovis* (Dworkin *et al.*, 2006). James Israel (1878) discovered a similar disease in humans, which was impossible to cultivate until Bujwid (1889) succeeded in isolating these organisms (Schaal *et al.*, 2006).

The Neighbor-Joining tree was constructed to display the relationship of members of the family Actinomycetaceae (Figure 1.17). The clusters were defined on the basis of 16S rRNA gene sequence similarities. Cluster 1 has core cluster 1, the members of which share 94-99% sequence similarity and are considered to be authentic *Actinomyces* species. *A. bovis*, *A. urogenitalis*, *A. israelii*, *A. gerencseriae*, *A. slackii*, *A. bowdenii*, *A. viscosus* and *A. naeslundii* were among the members of cluster 1, which were closely related to each other. Cluster 1A contained a distinct species of *A. graevenitzii* that share less than 94% sequence similarity with other species of the genus. Cluster 2 constituted *A. odontolyticus*, *A. meyeri* and *A. georgiae* which were closely related, while other species were moderately related. The species of Cluster 3, 4, 5 and 6 were phylogentically distinct from *Actinomyces* and placed in separate clusters. The main species constituting these clusters were of genus *Mobiluncus*, *Arcanobacterium* and *Actinobaculum* (Dworkin *et al.*, 2006; Schaal *et al.*, 2006).

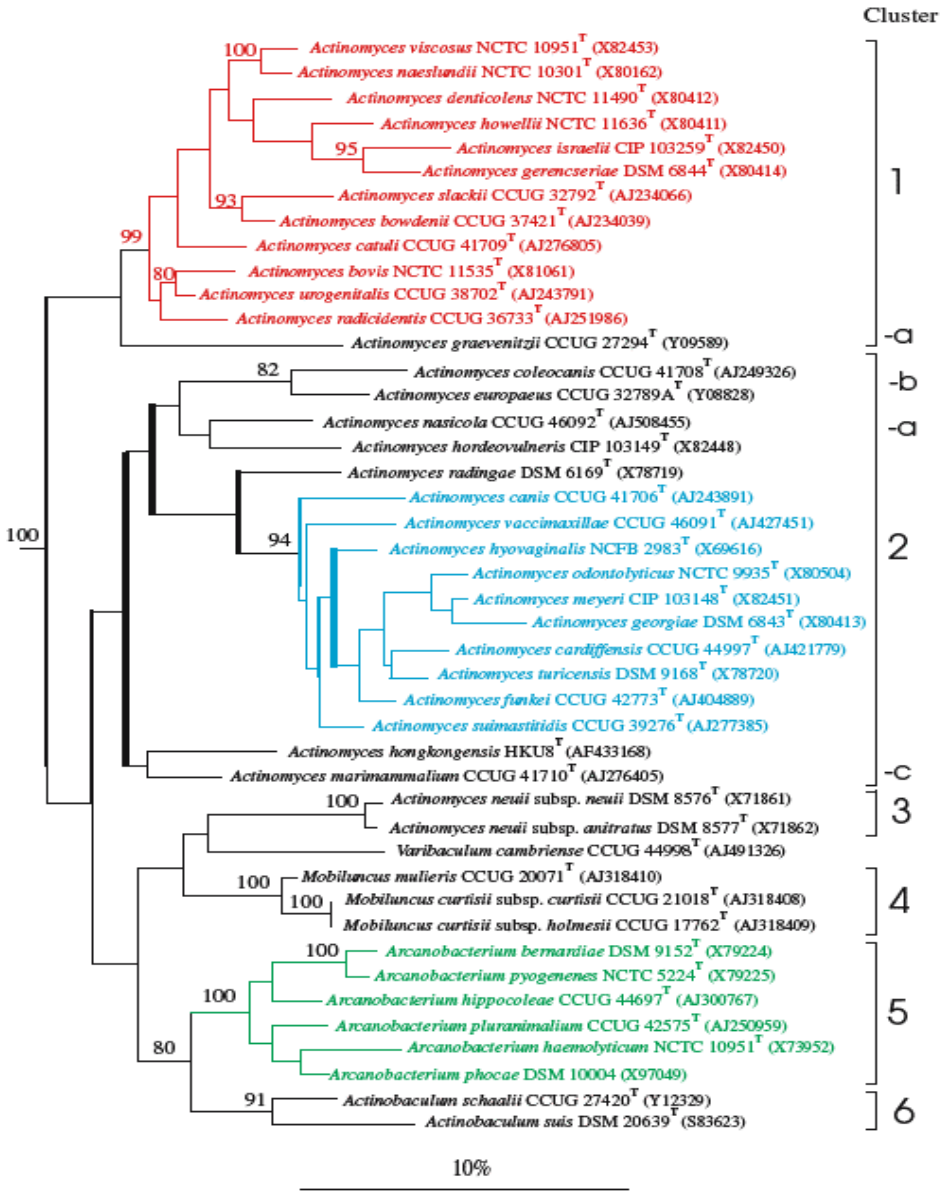


Figure 1.17: The discrimination of *Actinomyces* species into clusters based on 16S rRNA sequences of maximum-likelihood algorithm (Schaal, 2006).

1.6.2 Taxonomy Of Genus *Actinomyces*

Breed and Conn in 1919 first suggested that *Actinomyces* should be a genus name and *Actinomyces bovis* would be the type species (Winslow *et al.*, 1920). Before the report of Erikson (1940) it was very difficult to separate *A. bovis* from *A. israelii* due to a lack of suitable differential characteristics. In 1925 a filamentous bacterium was isolated by Carl Naeslund from the human oral cavity and subsequently in 1951 the name *A. naeslundii* was given by Thompson and Lovestedt (1951). *Actinomyces viscous* was also included in the genus *Actinomyces* after its definition changed to accommodate both catalase-positive and catalase-negative bacteria (Georg, 1969), previously it was under the genus designation "*Odontomyces*" and was isolated from hamsters (Howell *et al.*, 1965).

A. naeslundii and *A. viscous* were closely related to each other in antigenic and physiological characters (Fillery *et al.*, 1978; Holmberg & Forsum, 1973; Schaal, 1985; Schofield & Schaal, 1981). Gerencser (1979) suggested that *A. naeslundii* and *A. viscous* were varieties of a single species. A study by Schofield & Schaal (1981) revealed that strains labeled *A. naeslundii* and *A. viscous* formed a set of comparatively stable sub clusters. Almost similar results appeared in a recent evaluation of a new phenotypic differentiation system, which gave rise to four distinct numerical phenetic clusters that could be described within the *A. naeslundii* and *A. viscous* complex (Dahlen, 2004). These results can be confirmed at least in part by using molecular genetic analyses (Coykendall & Munzenmaier, 1979). In particular, the study of Johnson (1990) showed that strains belonging to *A. naeslundii* and *A. viscous* complex could be assigned to four genospecies based on DNA-DNA homology (Table 1.2). Four genospecies were "*A. naeslundii*, genospecies 1", "*A. naeslundii*, genospecies 2", "*Actinomyces* serotype WVA 963" and "*A. viscous*, serotype 1". *A. naeslundii* genospecies 1 contained species which were previously classified as serotype 1 and *A. naeslundii* genospecies 2 contained species previously known as serotype II, III, "*A. viscous* serotype II", and strains that can react with antisera of both strains of "*A. naeslundii*" and "*A. viscous* serotype II" known as serotype NV strains. "*Actinomyces* serotype WVA 963" formerly grouped under "*A. naeslundii* serotype IV" but was found to be distantly related to other organisms of this

complex that is why it was grouped as a separate genospecies and the same happened with the hamster isolate of "*A. viscous* serotype 1" that also formed a distant genospecies (Schofield & Schaal, 1981).

The Human Oral Microbiome Database (HOMD) project was launched and accessible through the web (www.homd.org) to collect 16S rRNA gene sequences and introduced them into a phylogeny-based database. HOMD is an annotated catalogue of 16S rRNA sequences derived from the amplification, cloning and sequencing of 16S sequences directly from oral samples and isolated bacteria. Researchers have classified many organisms based on 16S rRNA gene sequence analysis but this technique was not always reliable to classify bacterial species, especially viridans streptococci (Hoshino *et al.*, 2005), *Veillonella* species (Jumas-Bilak *et al.*, 2004) and lactobacilli (Naser *et al.*, 2007). Sequences of other genes including *sodA* (superoxide dismutase, Mn), *pheS* (phenylalanine tRNA synthetase, alpha subunit), *rpoA* (RNA polymerase, alpha subunit) *rpoB* (DNA-directed RNA polymerase, beta subunit) and *dnaK* (chaperone Hsp70, with co-chaperone DnaJ) were used to identify strains of these genera (Henssge *et al.*, 2009). A study by Tang *et al* (2003) also demonstrated that 16S rRNA gene sequences were not able to distinguish among the genospecies of *A. naeslundii* and other closely related species.

Unlabeled DNA from:	% Relatedness with reference DNA from:						
	A. naeslundii serotype II	*A. viscosus* serotype II	*Actinomyces* serotype NV[b]	*A. naeslundii* serotype III	*A. naeslundii* serotype I	*A. viscosus* serotype I	*Actinomyces* serotype WVA 963
A. naeslundii serotype II	100	62	63	51	33	24	30
A. viscosus serotype II	79	100	75	59	45	36	29
Actinomyces serotype NV	76[c]	70[c]	100	52[c]	41[c]	32[c]	35[c]
A. naeslundii serotype III	62	55	51	100	22	20	29
A. naeslundii serotype I	42	35	34	36	100	39	38
A. viscosus serotype I	44	36	35	37	55	100	36
Actinomyces serotype WVA 963	34[c]	33[c]	26[c]	31[c]	43[c]	33[c]	100

[a] Data were summarized from the data in Table 2. Values were determined by using the S_1 nuclease method.
[b] *Actinomyces* strains that react serologically with *A. viscosus* serotype II and *A. naeslundii* serotype I or II antisera.
[c] Values are the averages of the values shown in Table 2.

Table 1.2: The DNA relatedness data among genospecies of *Actinomyces* (taken from (Johnson *et al.*, 1990)

The recent study by Henssge *et al* (2009) investigated the phylogenetic relationship among selected *Actinomyces* strains and demonstrated that previously the designation used for genospecies of *Actinomyces* classified by Johnson (Johnson *et al.,* 1990) needed to be emended. Henssge *et al*., (2009) demonstrated that the genospecies could be differentiated based on determining the partial gene sequences of six house keeping genes *atpA* (ATP synthase F1, alpha subunit, ANA_0169), *rpoB* (DNA-directed RNA polymerase, beta subunit, ANA_1497), *pgi* (glucose-6-phosphate isomerase, ANA_0727), *metG* (methionyl-tRNAsynthase, ANA_1898), *gltA* (citrate synthase I, ANA_1674) and *gyrA* (DNA gyrase, subunit A, ANA_2224). *A. oris* that included *A. naeslundii* genospecies 2 previously known as serotype II, III, NV and the name *A. johnsonii* was given to *A. naeslundii* genospecies WVA 963 (Table 1.3).

Previous designation			Current designation
A. viscous	Serotype I [animal]	*A. viscous*	*A. viscous*
	Serotype II	*A. naeslundii*	*A. oris*
A. naeslundii	Serotype I	*A. naeslundii* [genospecies 1]	*A. naeslundii*
	Serotype II Serotype III Serotype NV Serotype (-)	*A. naeslundii* [genospecies 2]	*A. oris*
	Serotype WVA 963	*A. naeslundii* [genospecies WVA 963]	*A. johnsonii*

Table 1.3: Recent changes to taxonomy of *A. viscous*/*naeslundii*

The neighbour-joining tree was constructed using 7 concatenated genes, the six previously mentioned plus *pheS*, to determine the relationship between isolates identified as *A. oris* and *A. naeslundii*. It was found that both *A. oris* and *A. naeslundii* formed distinct clusters (Figure 1.18) supported by significant bootstrap values.

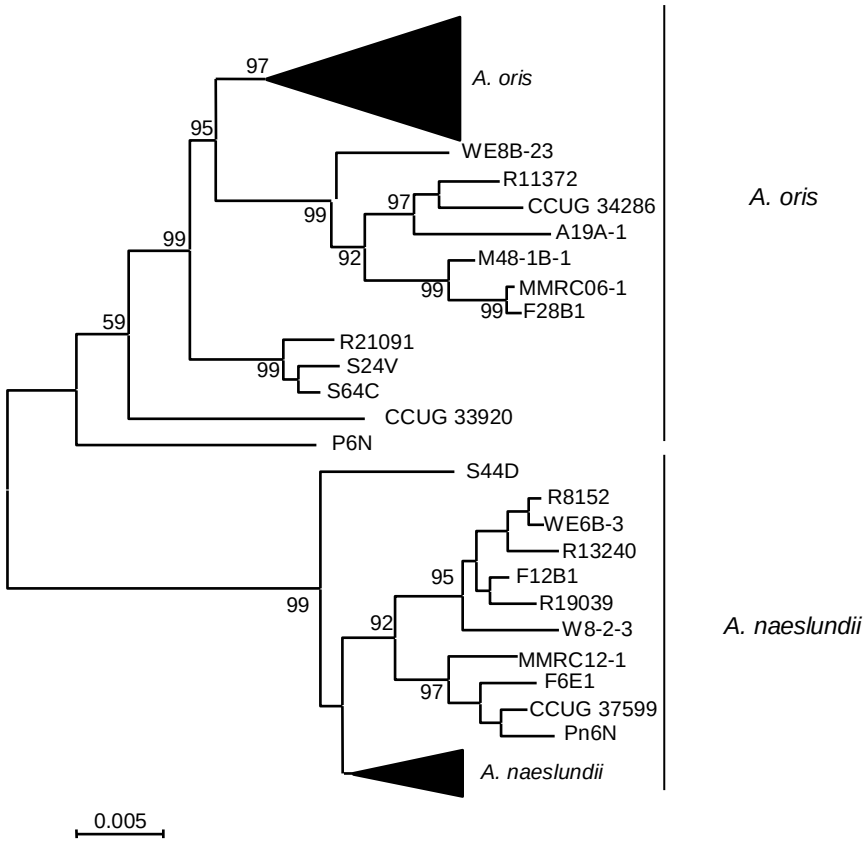


Figure 1.18: Neighbour-Joining tree of *A. naeslundii* and *A. oris* isolates using 7 concatenated house keeping genes. Bootstrap values are shown (taken with modification (Henssge *et al.*, 2011))

1.7 Lacto-*N*-Biose Operon

The whole genome sequence data obtained from the present study resulted in the identification of a lacto-*N*-biose operon in only the *A. oris* isolates. The lacto-*N*-biose metabolic pathway identified in *A. oris* is similar to the lacto-*N*-biose metabolic pathway in bifidobacteria (Beighton, 2013; Kitaoka *et al*., 2005). The important aspect of this finding is an improved understanding of the interaction of glycoprotein components of saliva with *A. oris* and its response specifically to the exposure to salivary mucins. Lacto-*N*-biose is a disaccharide derived from the O-linked glycans of mucins and glycoproteins. There is little known about the ability of oral bacteria to utilize these sugars although the metabolism of these sugars was extensively studied in relation to the influence of human breast milk on the infant gut flora and a novel pathway for the utilization of lacto-*N*-bioside has been described in *B. longum* NCC2705 (Kitaoka *et al*., 2005). The ability of bifidobacteria to utilize lacto-*N*-biose for growth may underpin the proliferation of these bacteria in the digestive tract of breast fed infants. It may therefore be expected that the oral bacteria with this ability would have an advantage over other species to proliferate more in the oral environment.

1.7.1 Lacto-*N*-biose Operon in bifidobacteria

Bifidobacteria are an important group of micro-organisms that reside in the human intestinal tract. They are Gram-positive, anaerobic, non-motile, non-spore forming, high GC content (55 to 67%) bacteria (Chen *et al.*, 2013; Do *et al.*, 2008; Do *et al.*, 2010; Murray *et al.*, 1992; Ventura *et al.*, 2004). There have been different studies which reported the presence of *Bifidobacterium* in the intestine of breast-fed infants (Bezirtzoglou *et al.*, 2006; Haarman & Knol, 2005; Harmsen *et al.*, 2000; Penders *et al.*, 2006; Yoshioka *et al.*, 1983). A study by Bezirtzoglou in (2006) reported the evidence of presence of *Bifidobacterium* in the intestine of new born infants after just 4 days following birth. The incidence of *Bifidobacterium* in male infants was greater as compared to female infants. A study comparing human milk with the milk obtained from

other sources recognized that human milk was unique in its features by containing complex oligosaccharides. The core tetrasaccharides in Human milk were Type I (lacto-*N*-tetraose) Galβ1-3GlcNAcβ1-3Galβ1-4Glc (Ito *et al.*, 2013). The oligosaccharides were one of the main components of human milk that were utilized by the group of *Bifidobacterium* (Coppa *et al.*, 2004), and was confirmed by the analysis of faecal samples in infants using FISH techniques. The prebiotic effect of human milk oligosaccharides was also verified (Satoh *et al.*, 2013). Human milk constituted 12-13 g/L of oligosaccharides. The chemical structure of more than 100 human milk oligosaccharides has been elucidated so far. The predominant oligosaccharides consisted of Lacto-*N*-Biose (LNB) type I (Gal (β1-3)-GlcNAc) structure. Type I LNB predominance is important because it may act as a substrate for beneficial bifidobacteria found in the gut of breast-fed infants (Urashima *et al.*, 2012).

The whole genome of *B. longum* NCC2705 was studied in detail and a putative Lacto-*N*-biose operon was found as shown in Figure 1.19. The operon consisted of a set of seven genes BL_1638 to BL_1644. BL_ 1638, BL_1639, BL_1640 are the components of ABC transporter proteins in *B. longum* and these proteins have a role to transport the oligosaccharide components from the extracellular environment to intracellular environment. Galacto-*N*-biose Phosphorylase (GLNBP, BL_1641) is an enzyme involved in the cleavage of human milk oligosaccharides or mucins. N-acetylhexosamine-1-Kinase (NahK, BL_1642); UDP-Glc/Gal1P Transferase (GalT, BL_1643) and UDP-Glc 4-epimerase (GallE, BL_1644) are the enzymes involved in utilization of Galacto-*N*-Biose from human milk oligosaccharides (Fujita *et al.*, 2005; Garrido *et al.*, 2012; Kitaoka *et al.*, 2005; Kitaoka, 2012; Yoshida *et al.*, 2012).

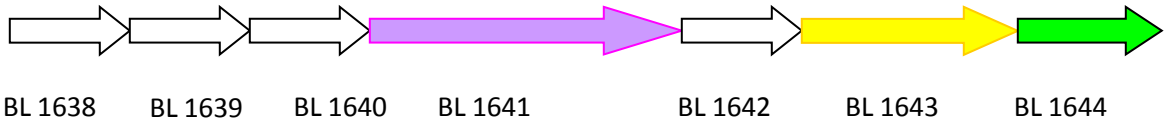


Figure 1.19: Putative Lacto-N-Biose operon found in *Bifidobacterium longum* NCC 2705 (taken from (Kitaoka *et al.*, 2005)).

Bifidobacterium bifidum inhabits the intestinal tract of breast-fed infants and is proposed to protect infants against diarrhoea and attack by pathogenic bacteria (Derensy-Dron *et al.*, 1999). Similarly Lacto-*N*-Biose Phosphorylase, an enzyme found and purified from the cell free extract of *Bifidobacterium bifidum* and the gene was found to be analogus to the BL_1641 gene found in *B. longum* NCC 2705 (Kitaoka *et al.*, 2005) that has the ability to degrade milk mucins. The enzymatic reaction involved the phosphorolysis of β-D-Gal (1,3)-D-GlcNAc (β-1-3-Galacto oligosaccharides) and converted into α-galactoside derivatives (Derensy-Dron *et al.*, 1999; Wada *et al.*, 2008).

β-D-Gal (1,3)-D-GlcNAc+Pi ⟶ α-D-Gal-1-*P* + GlcNAc

Figures 1.20 and 1.21 represents the metabolic pathway involved in the utilization of human milk oligosaccharides by the group of micro-organisms called bifidobacteria. Milk oligosaccharides contained Type 1 chain (Fucα1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glc) of Lacto-*N*-Biose (Kitaoka, 2012).

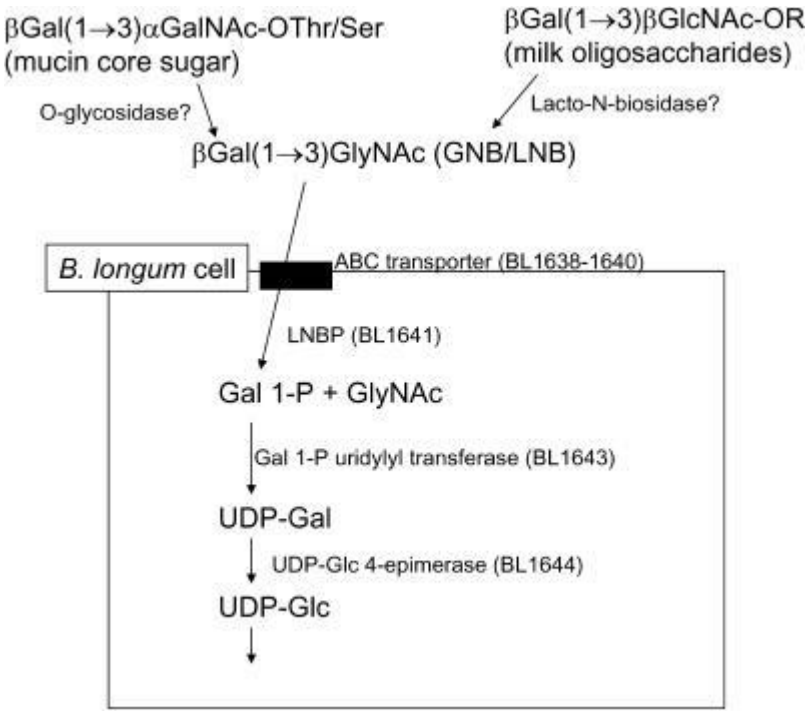


Figure 1.20: The proposed model of the Lacto-*N*-Biose metabolic pathway in *B. longum* (Kitaoka *et al.*, 2005).

1.7.2 Biological Functions of Oligosaccharides In Milk

The research on biological functions of human milk oligosaccharides has received much attention in the last century, with reports of having less gastrointestinal problems like diarrhoea in breast-fed infants than in the infants who were fed on formula milk in the first year of their life (Kunz & Rudloff, 1993). There was also some recent evidence about the breast milk effect in controlling urinary tract infections in newborns. Human milk was known to contain oligosaccharides, therefore these oligosaccharides were considered to play a role as an inhibitor of bacterial adhesion to epithelial cells that is a first step in the infective process (Kunz *et al.*, 2000). The concept behind this is that oligosaccharides act as a soluble receptor analogs of epithelial cell surface carbohydrates. The most recent functions observed were that it helped in binding of selectins (cell

adhesion molecules), which in turn stimulate the inflammatory reactions. The structure of oligosaccharides played a vital role in its biological functions. The presence of *Bifidobacterium bifidum* in the intestinal flora of breast-fed infants correlates with the fact that human milk contained a growth factor for these micro-organisms known as gynolactose (a mixture of 10 oligosaccharides) containing GlcNAc (N-Acetyl glucosamine), a growth promoting "bifidus factor" (Polonovski & Montreuil, 1954).

1.7.3 Oligosaccharides As Non-specific Defence Mechanisms

The oligosaccharides and glyco-conjugates in human milk have been shown to have a role to inhibit pathogenic microorganism (Kunz & Rudloff, 1993; Patton, 1994; Schroten *et al.*, 1992). The glycoproteins (milk mucins, MUC-1 also called as PAS-O and episialin) of human milk may be associated in the defence against infectious diseases therefore the evidence of diarrhoea and other infectious diseases are less profound in breast-fed infants rather than in bottle-fed infants. The defense mechanisms were involved in the attachment of milk mucins to bacterial receptors, which ultimately adhered to epithelial linings of the intestinal tract and inhibited the invasion of bacterial pathogens. This is the mechanism milk mucins had inhibited the experimental rotavirus infection (Yolken *et al.*, 1992) which was the leading cause of mortality due to acute gastrointestinal infections. The MUC-1 was also thought to have significance as a breast-tumor antigen (Beatty *et al.*, 2001). The exact mechanism of diarrhoea infection has not been established, but several bacteria like *Escherichia coli*, *Helicobacter jejuni*, *Shigella* strains, *Vibrio cholerae* and *Salmonella* spp. have the ability to adhere to the mucosal surface and cause infections in the small intestine or duodenum.

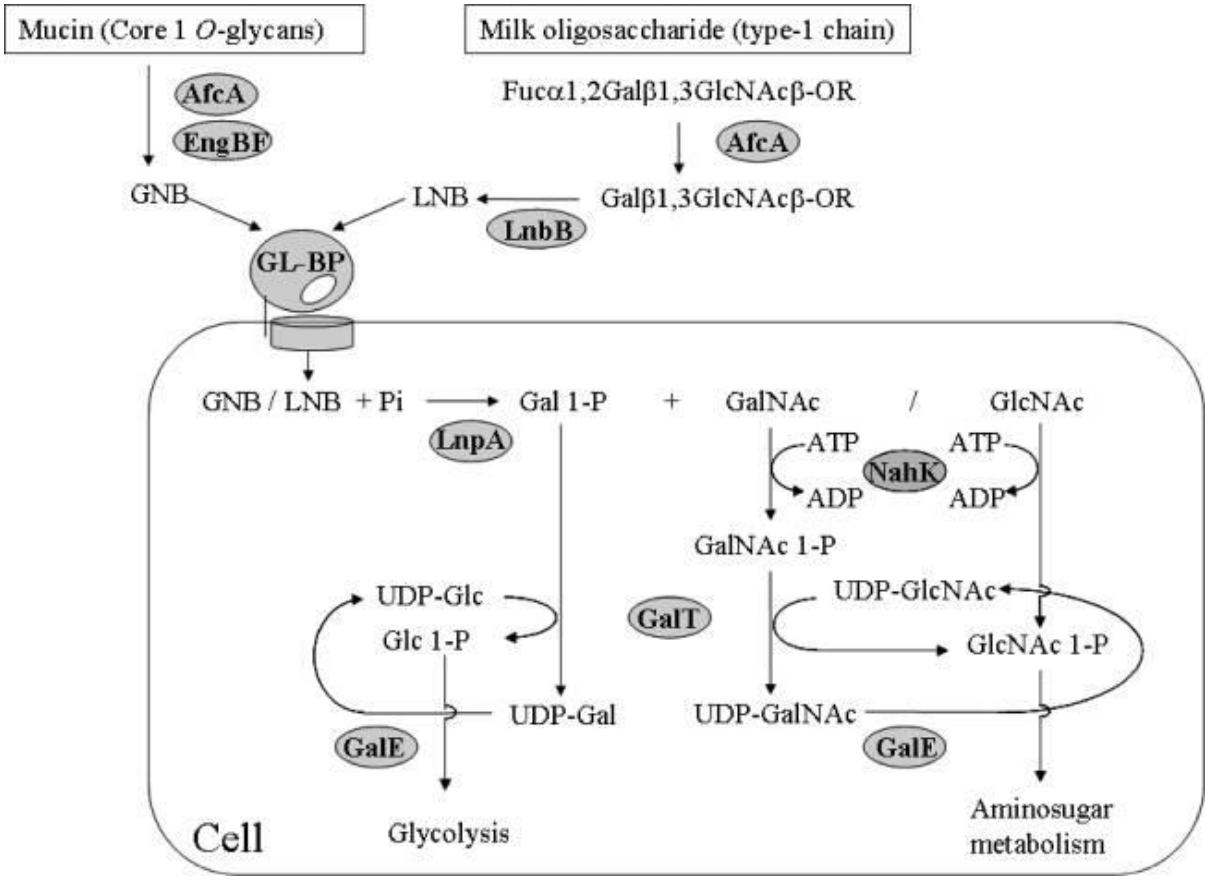


Figure 1.21: The metabolic pathway to utilize lacto/galacto-N-Biose from human milk oligosaccharides and mucin (Wada et al., 2008).

EngBF: extracellular endo-α-N-acetylgalactosaminidase, LnbB: lacto-N-biosidase, LnpA: lacto-N-biose phosphorylase, NahK: N-acetylhexosamine 1-kinase, GalT: UDP-glucose-hexose 1-phosphate uridylyltransferase, GalE: UDP-galactose epimerase.

1.8 Aims

The first hypothesis of the current research is that *A. oris* and *A. naeslundii* are the two main groups of actinomyces found in dental plaque and *A. oris* is better able to utilize carbon sources especially lacto-*N*-biose from the mucins therefore is better adapted to the oral environment than *A. naeslundii*. The second hypothesis is that *A. oris* and *A. naeslundii* both represents homogeneous group of strains.

The aim of this work was to investigate whether sub-clusters identified within *A. oris* and *A. naeslundii* by sequence analysis of the concatenated sequences of 7 housekeeping genes should be classified as sub-species or distinct species. For this purpose near-complete genome sequencing of selected strains of *A. oris* and *A. naeslundii* will be investigated. The ability of *A. oris* and *A. naeslundii* to utilize lacto-*N*-biose will be investigated in detail. Further, the ClonalFrame analysis will be used to investigate the phylogenetic status of the selected strains of *Actinomyces*.

Chapter 2 MATERIAL AND METHODS

2.1 DNA Extraction From *Actinomyces*

2.1.1 Bacterial Strains And Growth Conditions

The bacterial strains used in Next Generation Sequencing (NGS) study are listed in Table 2.1. These strains formed two outgroups in Figure 1.18. Frozen stocks were grown on fastidious anaerobic agar plates (FAA, LabM Ltd., Bury, Lancashire, UK) supplemented with 5% (v/v) defibrinated horse blood (TCS Biosciences Ltd., Botolph Claydon, Buckingham, UK) at 37 ºC anaerobically for 48 h. The purity of strains was verified by repeated streaking, microscopic examination of gram-stained cells and sequencing of 16S rRNA genes.

2.1.2 Lysis Treatment

A sterile 5 µl inoculating loop (Microspec Ltd., Cheshire, UK) was used to remove the cells (2-3 loops full) from the agar plate and suspended in 1ml of sterile water. The cells were washed by centrifugation (Microlite, Thermoelectron Corporation) at 13000 rpm for 5 min and the cell pellet resuspended in 500µl of TES buffer [(0.1M NaCl, 10mM Tris HCl and 1mM EDTA (pH:8.0, Sigma) and 5% (v/v) Triton (X-100, Sigma)]. The cells were mixed well and kept for 10 min at 37 ºC with mixing every 5 min. 100 µl of a solution of achromopeptidase from *Achromobacter lyticus* (5mg/ml, Sigma) and chicken egg white lysozyme (15mg/ml, Sigma) in TE buffer and 4µl of RNase solution (10mg/ml, Sigma) were added to lyse the cell wall and to digest the contaminating RNA. The suspension was incubated at 37 ºC for 2h. 50 µl of Proteinase K from *Tritirachium albam* (10mg/ml, Sigma), 10 µl of Pronase E (20mg/ml in TE buffer, Sigma) and 100 µl of 20% Sarkosyl (N-Lauryl sarcosine, Sigma) were added to complete cell lysis, to digest cell debris and remove extracellular polysaccharides and the cells were incubated at 37 ºC for 2h.

Table 2.1: Strains of *A. oris* and *A. naeslundii* included in this study

Name	Species	Reference/source	Origin
A19A-1	AO	Kings College London	leathery lesion
W11-1-1	AO	Kings College London	Plaque (caries +)
A7A-1	AO	Kings College London	Leathery lesion
G53E	AO	Kings College London	Plaque (caries -)
S64C	AO	Kings College London	Plaque (caries -)
F28B1	AO	Kings College London	Plaque (caries +)
M48-1B-1	AO	Kings college London	plaque caries
R11372	AO	Cardiff	IUCD
CCUG 33920	AO	P11N	Plaque
WE8B-23	AO	Kings CollegeLondon	Plaque (Caries +)
R21091	AO	Cardiff	Cerebral abscess
P6N	AO	Stromberg	Plaque
S24V	AO	Kings College London	Plaque (caries -)
F4D1	AO	Kings College London	Plaque (caries -)
R23275	AO	Cardiff	Blood culture
MMRCO6-1	AO	Kings College London	Soft lesion
CCUG 34286	AO	VPI D163E-3	Gingival crevice
Pn6N	AN	Stromberg	Plaque
W8-2-3	AN	Kings College London	Plaque (caries +)
S65A	AN	Kings College London	Plaque (caries -)
MMRC12-1	AN	Kings college London	soft lesion
R13240	AN	Cardiff	subphrenic abscess
CCUG 35334	AN	TF11	Blood(endocarditis)
CCUG 37599	AN	R709-03041/97	Cerebrospinal fluid
F12B1	AN	Kings CollegeLondon	Plaque (Caries +)
WE6B-3	AN	Kings CollegeLondon	Plaque (Caries +)
NCTC 10301	AN	Stromberg	Plaque
S43L	AN	Kings College London	Plaque (caries -)
G127B	AN	Kings College London	Plaque (caries -)
MB-1	AN	Kings College London	Plaque (caries +)
R24330	AN	Cardiff	IUCD
T23P-1	AN	Kings College London	Plaque (caries -)
S44D	AN	Kings College London	Plaque (caries -)
F6E1	AN	Kings College London	Plaque (caries -)
R19039	AN	Cardiff	Liver abscess
R8152	AN	Cardiff	IUCD

* IUCD: Intra-uterine contraceptive devices ** AO: *A. oris*, AN: *A. naeslundii*

2.1.3 Purification And Isolation Of Genomic DNA

To the lysed cell suspension, 500 µl of saturated phenol:chloroform:isoamyl alcohol (25:24:1, v/v/v, Sigma) was added, to precipitate proteins and debris, and vortexed for a few seconds. The tubes were centrifuged at 13,000 rpm for 20 min. The upper layer containing DNA was removed and transferred into a new tube and the phenol:choloroform extraction was repeated until the upper aqueous layer was clear. The clear upper aqueous layer was transferred to a new tube and 60 µl of 3M (pH 5.0) sodium acetate and 1000 µl of 100% ethanol (at -20 °C) were added to precipitate the DNA. The tubes were left for 15 min at -20 °C. The tubes were centrifuged at 4 ºC again at 13,000 rpm for 20 min, the supernatant decanted, the pellet washed twice with 1000 µl of 70 % ethanol (at -20 °C) to dehydrate the DNA and rehydrated in 30 µl of TE buffer. The purified extracted DNA was stored at 4 °C for short term and at -20 °C for long term.

The other effective method which was used to isolate the DNA was by modifying the protocol of The GenElute^{TM} bacterial Genomic DNA (NA 2100, Sigma Aldrich) kit. The kit was used for the isolation of DNA after lysis treatment using the combination of enzymes as described above. The pre-assembled GenElute Miniprep Binding Column (with the red O-ring) was prepared with a 2ml collection tube and 500 µl of column preparation solution was added and centrifuged at 12,000 x g for 1 min. Elute was discarded. The column preparation solution was used to maximize the binding of DNA to the membrane to obtain consistent yield. 200 µl of ethanol (95-100%) was added to the above lysate and mixed by vortexing for 5-10 sec. The homogenous mixture is required for further stages of DNA isolation. The entire content of the tube was transferred onto the binding column. A wide bore pipette tip was used to reduce the shearing of DNA. The contents were centrifuged at 6500 x g for 1 min. The collection tube was discarded and the column was placed in a 2ml collection tube to which 500 µl of wash solution 1 was added and then centrifuged at 6500 x g for 1 min. The collection tube was discarded and replaced with a new collection tube. The column was washed with 500 µl of wash solution and centrifuged for 3 mins at 16,000 x g to dry the column. The column was dried until no residues were left on the column. The elution solution

was added directly onto the centre of the column and the column was centrifuged for 1 min at 6500 x g to elute the DNA. The elution efficiency was increased by again transferring to elute from the columns at the same centrifuge speed.

2.1.4 Quantification Of DNA Using Picogreen

The quantity of DNA was measured with the Quant-iT^{TM} dsDNA Assay Kit (Molecular Probes Inc., Eugene, USA) in a microtitre plate. 100 µl of a 1:100 dilution of the DNA extract were added to 100 µl of a 1:200 PicoGreen® solution. All dilutions were made with the TE buffer (pH 7.5) provided. Fluorescence was measured with an excitation wavelength of 480 nm and emission at 520 nm. A standard curve (Fig: 2.1) was prepared with the lambda DNA standard (100µg/ml, Sigma).

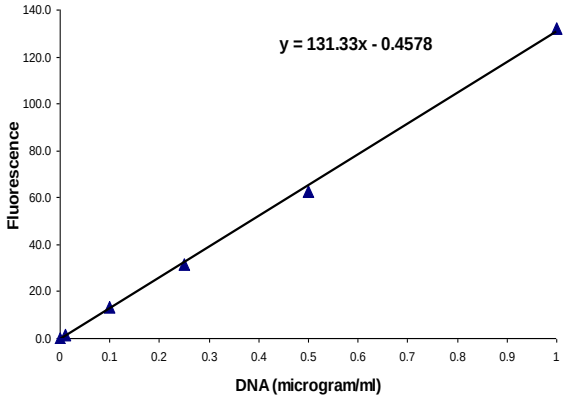


Figure 2.1: The standard curve

*** The concentration was shown of DNA in µg/ml on the X-axis and average DNA fluorescence on the Y axis.**

2.2 Library Preparation For Illumina Paired-End Sequencing

2.2.1 Paired-End (2x76bp) Multiplexed Sequences

The Paired-end multiplexing sequencing assay protocol was followed to generate high quality sequences from both ends of DNA inserts. The paired-end run enables DNA sequencing up to 2 x 76 bp reads for fragments ranging from 150-200bp and generate up to 200 million reads in single run. The raw multiplexed sequences were obtained from Illumina Analyzer and kept on a server.

2.2.2 DNA Quantification By Qubit

For the Qubit measurement of DNA concentration in the genomic DNA samples, the tubes were set up for the standards and samples. The reagents for qubit measurement were provided in Quant-iT dsDNA BR (broad range) assay kit (Invitrogen) used with Qubit fluorometer (Invitrogen). The Quant-iTTM working solution was prepared by diluting the Quant iT reagent 1:200 in Quant iT buffer (Table 2.2). The reading was multiplied by the dilution factor using the formula;

DNA concentration = QF value × 200/x

QF value was the reading given by Qubit fluorometer (shown in Figure 2.2) and x was the number of µl of sample added to the assay tube.

Table 2.2: The worksheet to prepare sample and standards to quantify dsDNA

	Standard Assay Tubes	User Sample Assay Tubes
Volume of Working Solution to add	190 µl	180-199 µl
Volume of Standard to add	10 µl	-
Volume of User Sample to add	-	1-20 µl
Total Volume in each Assay Tube	200 µl	200 µl

Figure 2.2: QubitTM Fluorometer and assay reagents
(figure taken from http://probes.invitrogen.com/media/publications/605.pdf)

2.2.3 DNA Shearing

DNA was sheared using the Covaris system S2 (Figure 2.3). Basically, 3µg of high-quality genomic DNA was diluted with 1xTE Buffer in a DNA LoBind tube (Eppendorf) to a final volume of 120 µl. The 120 µl of DNA sample was slowly transferred using tapered pipette tip through the pre-split septa into the Covaris microtube. The Covaris system settings (Table 2.3) were used as instructed in the manual to shear the DNA of fragments of 150±10%. The sheared DNA was purified using QIAquick PCR Purification Kit (Qiagen) and the quality of DNA was analyzed with Agilent 2100 Bioanalyzer (Agilent).

Figure 2.3: Covaris S2

Table 2.3: Covaris system settings

Setting	Value
Duty Cycle	10
Intensity	5
Cycles per Burst	200
Time	6 cycles of 60 seconds each
Set Mode	Frequency sweeping
Temperature	4 ℃

2.2.4 Assess The Quality Of DNA With Agilent 2100 Bioanalyzer

The quality of the DNA was analyzed with Agilent 2100 Bioanalyzer (Agilent p/n G2938C). The chip used was Agilent "DNA 1000" along with the reagent kit (Agilent DNA 1000 kit, Agilent p/n 5067-1504). The instrument electrodes were checked before the start of the experiment and cleaned as instructed in the reagent kit guide. The Agilent 2100 expert software (version B.02.02 or higher) was opened and the correct programme was selected and communication was checked with the hardware. The chip, samples and ladder were prepared as instructed in the reagent kit guide. The prepared chip was loaded into the Agilent 2100 Bioanalyzer and runs were started within five minutes after preparation of the chip. The appropriate assay was chosen from the drop down list within the instrument context. The run was started. The sample name and comments were

entered in the data and assay context. The results were verified by checking the electropherogram peak which shows the distribution with the peak height of 150 ± 10% nucleotides (Figure 2.4).

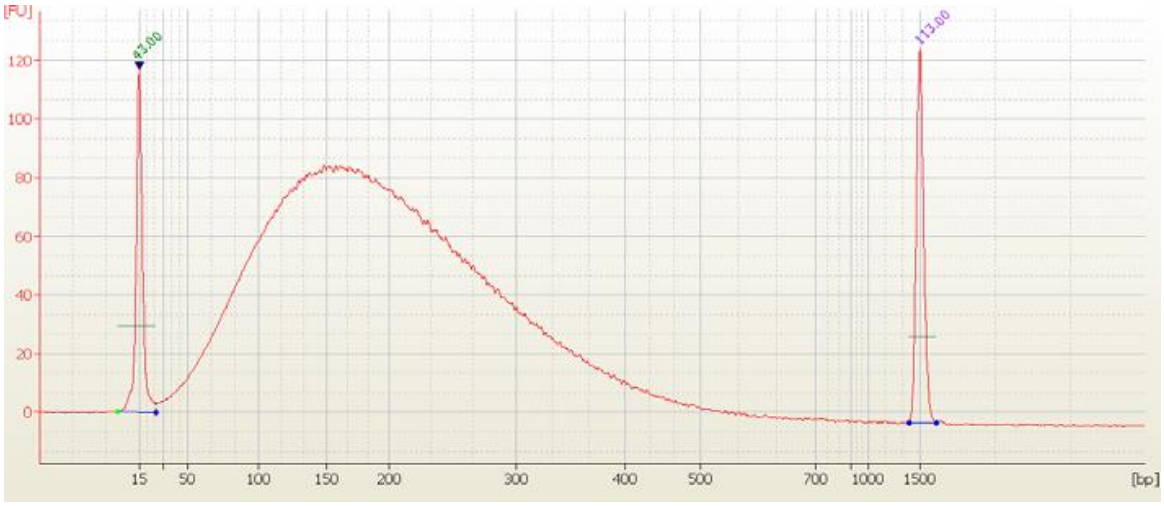


Figure 2.4: Electropherogram showing peak height of DNA fragments between 150-200bp

2.2.5 Repair The Ends

The DNA ends were repaired as blunt ends were created during fragmentation of DNA. The enzymes used to repair the ends were T4 DNA polymerase and DNA polymerase I Klenow fragment. The 3' to 5' exonuclease activity of these enzymes was used to cut the 3' ends (blunt ends) and polymerase activity was used to fill in with the 5'-phosphorylated end. The kit used to repair the ends was NEB (New England Biolabs, E6000, UK, Ltd). The reagents to repair the ends were used according to instruction given in the manual and listed in Table 2.4. The reactions were mixed well by gently pipetting and incubated for 30 mins at 20 °C. The DNA was purified as previously described.

Table 2.4: The reagents used to repair the blunt ends with 5'-phosphorylated ends

Reagent	Volume for 1 Library
DNA sample	29 µl
Nuclease-free water	46 µl
10X Phosphatase Buffer with 10mM ATP	10 µl
DNTP solution mix with 10mM	4 µl
T4 DNA Polymerase	5 µl
DNA Polymerase I,Large (Klenow) fragment enzyme	1 µl
T4 Polynucleotide Kinase (PNK)	5 µl
Total Volume	100 µl

2.2.6 Add 'A' Bases To The 3' End Of The DNA Fragments (A-Tailing)

'A' bases were added to the 3' end of the blunt phosphorylated DNA fragment using the polymerase activity of the Klenow fragment. This step specifically prepared the DNA fragment for ligation to adapters, which have a single 'T' base overhang at the 3' end. The reaction mix was prepared using reagents given in protocol for each library and listed in Table 2.5. The samples were incubated for 30 mins at 37 °C. The DNA was purified using MinElute PCR purification kit (Qiagen).

Table 2.5: The reagents used to add 'A' bases to the template DNA

Reagent	Volume
DNA Sample	32 µl
10X NEBuffer 2	5 µl
dATP Solution with 1mM	10 µl
Klenow exo (3' to 5' exo minus) or (3' - > 5' exo -)	3 µl
Total Volume	50 µl

2.2.7 Ligation Of Paired-End Adapter

Adapters were ligated to the ends of the DNA fragments so that they would be ready to hybridize to flow cells. The index-specific adapters were ligated to prepare adapter-modified DNA ends. The reaction mix was prepared using reagents in the manual and listed in Table 2.6 and was kept on ice. The samples were incubated for 15 minutes at 20 °C. The SPRI bead purification system was used for the removal of any unligated adapters and quality was assessed with Agilent 2100 Bioanalyzer.

Table 2.6: Reagents to ligate the adapters at the ends of DNA fragment

Reagent	**Volume of 1 library**
DNA sample	10 µl
Nuclease-free water	4 µl
2X Quick Ligation Reaction Buffer	25 µl
Paired-End Adapter oligo mix (illumina kit)	6 µl
Quick T4 DNA ligase	5 µl
Total Volume	**50 µl**

2.2.8 Quality Assessment with Agilent 2100 Bioanalyzer

The quality of ligated DNA fragments were assessed with the Agilent 2100 Bioanalyzer using High sensitivity DNA assay and the procedure was carried out as described previously in this protocol. The results were verified by seeing a shift in the peak height of the size of the DNA fragments upto 200 to 250bp (Figure 2.5) as were found after shearing DNA.

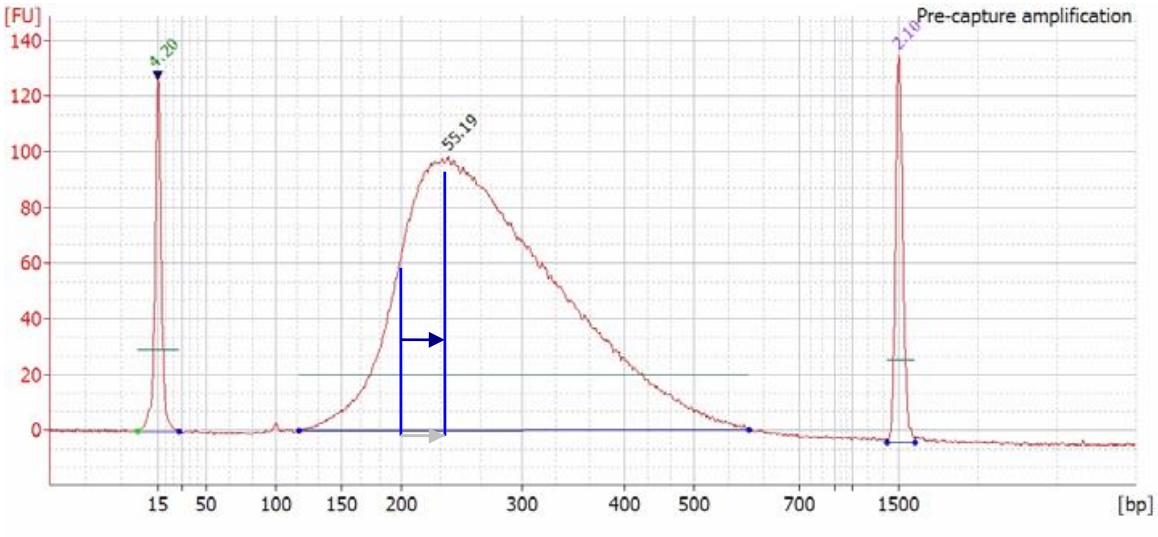


Figure 2.5: Ligated Library assessments by looking at the shift of peak height (blue arrow) from 200 to 250bp.

2.2.9 Amplification Of Adapter-ligated Library By PCR

The DNA fragments which had adapters at both ends were amplified and the six-base tag was added in this step. The InPE 1.0 and InPE 2.0 primers were used along with index primer (12 primers in a set) and reaction products were put in the PCR machine and during this stage primers will anneal to the ends of adapters. The reaction mix was prepared using the reagents from Multiplexing Sample Preparation Oligonucleotide kit (Illumina) according to manufacturer's instructions and reagents used were listed in Table 2.7, in 200 µl thin walled PCR tubes using the appropriate PCR primer index for each separate sample. The amplification conditions were set according to the following PCR programme and purified using AmPure XP beads. The library was finally validated with the Agilent 2100 Bioanalyzer using the High Sensitivity Chip. The PCR programme used is as follows (a) 98 ºC for 30sec (b) 98 ºC for 10 sec, (c) 65 ºC for 30sec (d)72 ºC for 30sec, go to (b) 18 times (e)72 ºC for 5min and hold at 4 ºC

Table 2.7: Reagents to enrich the adapter-modified DNA fragments by PCR

Reagent	**Volume (µl)**
DNA	3
DMSO	1.5
Ultra pure water	To increase DNA volume to 17.5
PCR primer InPE 2.0	1
PCR primer InPE 1.0	1
PCR primer index [#]	1
*Phusion DNA polymerase (Finnzymes Oy)	25
Total volume	**50**

***25 µl of Phusion High-Fidelity PCR Master Mix with HF buffer (Cat No: F-531L, Finnzymes Oy, Finland) was added in a total volume of 50 µl.**

2.2.10 Cluster Generation

The libraries were prepared for generating clusters using the cBot/cluster station. Briefly 15 µl of 1nM library was prepared by diluting the stock library concentration with nuclease free water. PhiX neat concentration (Illumina) was 10nM. 1.6 µl of 10nM stock PhiX was diluted with 14.4 µl of distilled water to give 15 µl of 1nM PhiX. The 10 µl HP3 (2M NaOH supplied with cluster generation kit) was diluted with 190 µl of elution buffer. The 16 µl of diluted HP3 was added to 16 µl of diluted libraries. The libraries were incubated for 5 min at room temperature to denature and then immediately 968 µl of chilled buffer HT1 was added to 30 µl of denatured library and vortexed briefly and finally kept on ice. Aliquots (120 µl) of each library were added in strip tubes. When 16pM library was ready to load into flow cell (Figure 2.6), the cBot was run for 4h. After generating clusters the flow cell was loaded onto Illumina Genome Analyzer$_{IIxe}$.

Figure 2.6: Flow cell showing eight lanes on single slide.

2.3 Whole Genome Sequencing Using Genome Sequencer By 454 Life Sciences

The Genome Sequencer (GS) 454 (Branford, CT, USA) was used for the preparation of the genomic library of strains listed in Table 2.1. The principle of GS of 454 was related to using PCR-based techniques. The system was not suitable for *de novo* sequencing initially due to only 20-30 base sequences could be obtained but with the advent of the advanced version, it was possible to get up to 450 bases with the GS (FLX) platform using Titanium reagents. The GS (FLX) platform was able to produce millions of sequences per run that gave coverage of around 6 x (on a 4Mb genome).

2.3.1 Single-End Sequencing Using Roche 454 GS (FLX) Titanium Platform

The sample requirement for Rapid library preparation for Roche 454 GS (FLX) titanium analyzer was the double stranded DNA with an $OD_{260/280} \geq 1.8$ with a concentration of 500 ng and fragment size of more than 1.5kb. The Single-end sequencing assay protocol of Roche 454 was followed on Roche GS (FLX) analyzer to generate long insert reads of about 400bp. High quality sequences from Single end of the same DNA inserts were obtained.

2.3.2 Multiplex Identifier (MID) Adaptors For Rapid Library Preparation

The GS (FLX) Titanium sequencer can run 12 Multiplex Identifier (MID) Adaptors (numbers 1-12) that are available in the current GS (FLX) Titanium Rapid Library MID Adaptors Kit (Part No. 05619211001). The final construct of DNA with MID adapter is shown in Figure 2.7 and the list of MID sequences used is mentioned in Table 2.8.

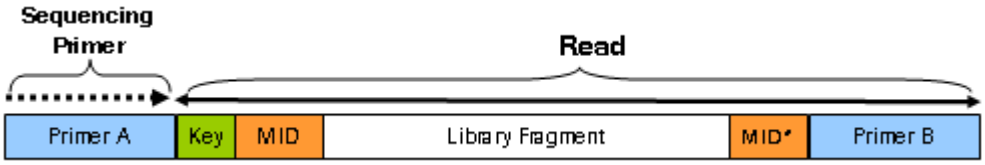


Figure 2.7: Final library constructs with MID adapters

Table 2.8: List of MID sequences

1	ACACGACGACT
2	ACACGTAGTAT
3	ACACTACTCGT
4	ACGACACGTAT
5	ACGAGTAGACT
6	ACGCGTCTAGT
7	ACGTACACACT
8	ACGTACTGTGT
9	ACGTAGATCGT
10	ACTACGTCTCT
11	ACTATACGAGT
12	ACTCGCGTCGT

*** Sequences were obained from Technical Bulletin of Genome Sequencer FLX system**
***(**http://www.liv.ac.uk/media/livacuk/centreforgenomicresearch/The_GS_FLX_Titanium_Chemistry_Extended_MID_Set.pdf)

The protocol used was a modified version of the GS-FLX-Titanium-Rapid-Library-Preparation-Method-Manual and NEBNext Quick DNA sample Prep Set2 (E6080S) to prepare the whole genomic DNA library. 500 ng of extracted genomic DNA was taken in a 1.7 ml centrifuge tube. TE buffer was added to a final volume of 130 µl. The DNA sample was fragmented using Covaris^{TM} S2 system with the 400bp settings. The settings used were 10% Duty cycle; Intensity was 4, 200 cycles per burst and run for 55 sec. The fragmented DNA was transferred to a fresh eppendorf tube.

Qiagen MinElute PCR purification kit was used to purify the fragmented DNA sample. 650 µl PBI buffer was added in 130 µl of sheared DNA sample. The contents were mixed by pipetting up and down. 750 µl of sample was loaded onto the column. The DNA sample was centrifuged for 15 sec and flow-through was discarded. 750 µl of PE

buffer was added and the column was centrifuged again for 1 min and flow-through was discarded. 16 µl of TE buffer was added to elute the DNA and centrifuged in a fresh tube. The sample was transferred to a 200 µl PCR tube.

2.3.3 Fragment End Repairs

The following (Table 2.9) end repair mix was prepared in a 2.0 ml centrifuge tube.

Table 2.9: Preparation of End Repair Mix

Reagents	**Volume (µl)**
RL 10 × PNK Buffer	2.5 µl
RL ATP	2.5 µl
RL dNTP	1 µl
RL T4 Polymerase	1 µl
RL PNK	1 µl
RL Taq Polymerase	1 µl
Total volume	**9 µl**

The 9 µl of End repair mix was added to a DNA sample. The contents were vortexed for 5 sec and spun for 2 sec in a mini centrifuge (Microlite RF, Thermo Electron Corporation). The samples were run on a thermocycler using the PCR programme at 25 ºC for 20 min, 72 ºC for 20 min, and 4 ºC on hold with the heated lid on.

2.3.4 AMPure Bead Preparation

Agencourte AMPure Beads were prepared to clean up the contents after running on the thermocycler. The AMPure bead bottle was vortexed for 20 sec. 125 µl of AMPure beads were aliquot in 200 µl PCR tubes. The tubes were placed on a Magnetic Particle Concentrator (MPC). The beads were collected on the side of the tube and the

supernatant was carefully removed and discarded. Sizing solution (125 µl) was added and vortexed for 5 sec.

2.3.5 Adapter Ligation

1 µl of RL MID adapter was added in reaction mix. 1 µl of RL Ligase was added to the reaction mix. Reaction mix was vortexed for 5 sec and centrifuged for 2 sec in mini centrifuge. The reaction mix was incubated for 10 min at 25 ºC.

2.3.6 Small Fragment Removal

The samples were added to the AMPure beads prepared in sec 2.3.4. Samples were vortexed for 5 sec and centrifuged for 2 sec. The tubes were incubated for 5 mins at room temperature. The tubes were placed on the Magnetic stand. The beads were allowed to be collected on the walls of the tube and the supernatant was discarded. 25 µl of TE buffer was added in PCR tubes and vortexed for 5 sec. 500 µl of sizing solution was added and vortexed for 5 sec. The contents were incubated at room temperature for 5 min. The tubes were kept on a magnetic stand and beads were washed twice with 175 µl of 70% ethanol. The ethanol was removed completely. The beads were air dried at room temperature for 2 min. 53 µl of TE buffer was added to the beads and were left for 2 min on the magnetic stand. 50 µl of supernatant was transferred to the new tube.

2.3.7 Library Quality Assessment

1 µl of DNA aliquot was loaded on to a Agilent Bioanalyzer High sensitivity DNA chip to assess the quality according to the protocol mentioned in section 2.2.4.

2.3.8 Preparing Working Aliquots

The aliquot of DNA library was diluted to a working stock of 1 x 10^7 molecules /µl in TE buffer. The working stock was transferred into small tubes in aliquots of 25 µl and stored at -25 ºC before proceeding to emPCR amplification.

2.3.9 Library Quantification By KAPA Kits

The manufacturer's protocol was used to quantify the library and the kit used was KAPA Library Quantification Kit (KK4821, KAPA Biosystems, USA) for the Roche 454 GS Titanium platform. The qPCR/Primer mix was prepared by adding 1ml of 454 Titanium Primer Premix (10X) to 5 ml bottle of KAPA SYBR® FAST qPCR Master Mix (2X) and mix well. 1:500 dilution of dsDNA library was prepared by diluting the library DNA in 10 mM Tris-HCl, pH 8.0 + 0.05 % Tween 20. qPCR plate was prepared by adding 12 µl of KAPA SYBR® FAST qPCR Master Mix containing Primer Premix , 4 µl of PCR-grade water and 4 µl of diluted library DNA or DNA standard (1-6). The qPCR protocol was run at 95 ºC for 5 min for initial activation, and at 95 ºC for 30 sec for denaturation and 35 cycles of Annealing/extension/data acquisition at 60 ºC for 90 sec. Data was analyzed by confirming 90-100 % reaction efficiency for standards and library and then the library concentration was calculated. Serial dilutions were prepared of undiluted DNA library for entry into emPCR.

2.3.10 Emulsion PCR (emPCR) Amplification

2.3.10.1 Preparation Of Reagents And Emulsion Oil

The Emulsion PCR Reagents box was opened and the kit components were thawed at room temperature except for Enzymes Mix and Pipase tubes which were kept at -15 to -20 ºC. The tubes were vortexed for 5 sec. The tube of additive was heated at 55 ºC for 5 min to dissolve the contents properly. The remaining precipitate was centrifuged and the supernatant was used. All components of kits were centrifuged including the enzymes in a bench top mini centrifuge for 10 sec. The enzymes were returned to -15 ºC. The other reagents were kept at room temperature.

2.3.10.2 Preparation Of Mock Mix And Pre-emulsion

The tube of emulsion oil was vigorously vortexed for 10 sec at maximum speed and the entire contents were poured (4ml) into the Turrax stiring tube. 1 x Mock Mix was prepared by adding 430 µl of Mock Mix to 1.72 ml of molecular biology grade water and was vortexed to mix. The 2.0 ml of 1 x Mock Mix was added to the Turrax stirring tube containing the emulsion oil. The Ultra Turrax Tube Drive (UTTD) was set to 4000 rpm for 5 minute. The stirring tube was put into UTTD and UTTD was started to mix the emulsion.

2.3.10.3 Preparation Of The Live Amp Mixes A And B

The Live Amp Mix A and B was prepared according to Table 2.10 in two separate tubes. The contents were vortexed for 5 sec and stored on ice.

Table 2.10: Preparations of the Live Amp Mixes A and B

A: Live Amp Mix A

Reagent	**Volume (µl)**
Mol.Bio.Grade.water	205
Additive	260
Amp mix	135
Amp Primer A	40
Enzyme Mix	35
PPiase	1
Total	**676**

B: Live Amp Mix B

Reagent	**Volume (µl)**
Mol.Bio.Grade.Water	205
Additive	260
Amp Mix	135
Amp Primer B	40
Enzyme Mix	35
PPiase	1
Total	**676**

*** Amp Mix A and B:** Amplicon mixture A and B each have set of primers to be annealed at the end of DNA amplicon fragments**.**

2.3.11 DNA Library Capture

0.5 ml of 1 x wash buffer was mixed with 4.5 ml of Molecular Biology Grade Water. 1 ml of 1 x wash buffer was added in both tubes of capture beads and was vortexed. The capture beads were pelleted by spinning for 10 sec. The supernatants were discarded without disturbing the pellets. The capture beads were washed again with 1 ml wash buffer, vortexed, spun down and the supernatant was discarded. An aliquot of amplicon DNA library was thawed to be amplified. The volume of DNA library was measured by using the following equation.

$$\mu l \text{ of DNA library per tube} = \frac{\text{Desired molecules per bead x 5 million beads}}{\text{Library concentration (in molecules/}\mu l)}$$

The calculated volume of the amplicon DNA library was added in the tube of washed Capture beads A and in Captured Beads B tube separately. The tubes were vortexes for 5 sec.

2.3.12 Emulsification

The 600 µl of Live Amp Mix B was added to the tube of captured library A and vice versa. The contents were transferred into the Turrax stirring tube (Figure 2.8 and 2.9). The tubes were placed in UTTD and rotated at 2,000 rpm for 5 minutes.

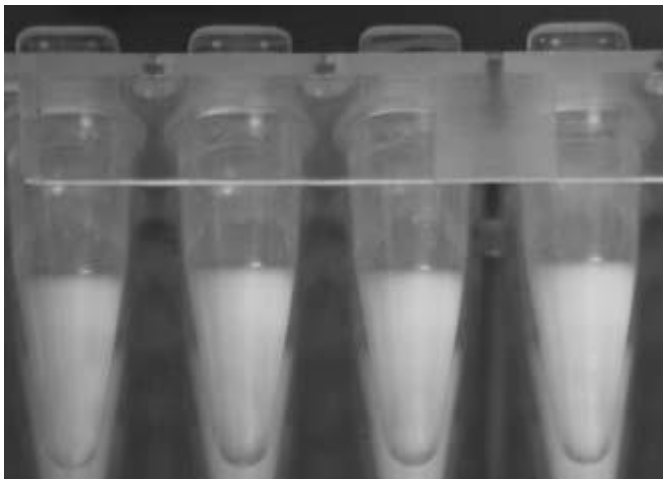


Figure 2.8: DNA fragments suspended in an emulsion mixture as a homogenous suspension

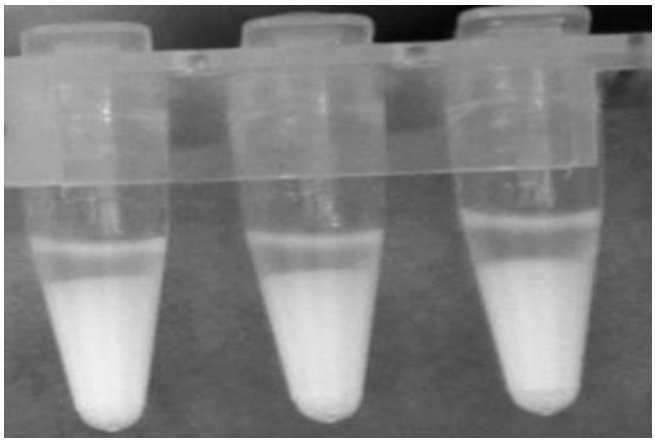


Figure 2.9: DNA fragments separates as a clear liquid in an emulsion mixture.

2.3.13 Amplification

2.3.13.1 Dispensing The Emulsions

100 µl of emulsion was added in the PCR tubes. The tubes were capped and sealed properly.

2.3.13.2 Amplification Reactions

PCR tubes were placed in the thermocycler with the heated lid on and the following PCR programme was run: 1 x 4 min at 94 ºC and then 50 x 30 sec at 94 ºC, 4.5 min at 58 ºC, 30 sec at 68 ºC, and finally hold at 10 ºC.

2.3.14 Bead Recovery

2.3.14.1 Vacuum-Assisted Emulsion Breaking Set-up

The GS Junior Titanium emPCR Oil and Breaking kit was brought to the external ventilated hood. 50ml of conical tube was attached to the lid from the GS Junior Titanium Oil and Breaking Kit. The blue connector was inserted into the top opening of the transpette. The other end of the tubing was connected to the vacuum source (Figure 2.10).

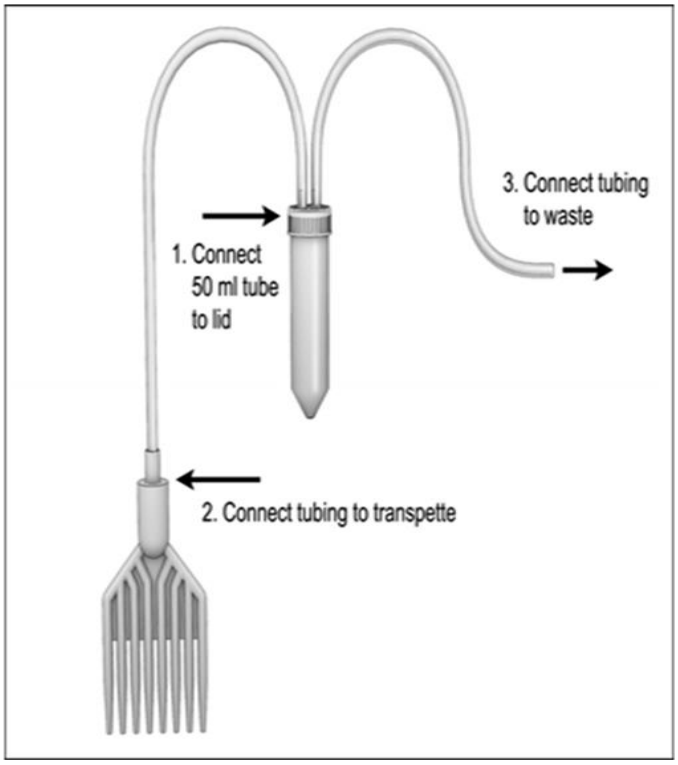


Figure 2.10: Assembled set up of vacuum assisted emulsion breaking and bead recovery system

***(adapted from emPCR Amplification method manual-Lib-A, May 2010).**

2.3.14.2 Emulsion Collection And Initial Washes

The vacuum was turned on and emulsion (A and B) were aspirated from the wells and were collected in the 50 ml tube using transpette tips at the bottom of the wells. The wells were rinsed with 100 µl of isopropanol per well. The vacuum was turned off and 50 ml tube was removed which contained beads with amplified DNA.

2.3.14.3 Bead Washes And Recovery

The 50 ml tube of collected emulsions was vortexed. Isopropanol was added to a final volume of 35 ml and vortexed to resuspend the pellets. The beads were centrifuged at 930 x *g* for 5 min and supernatant was poured out. 10 ml of enhancing buffer was added and vortexed thoroughly to resuspend the pellet. The washing steps were repeated twice. The DNA bead suspension was transferred using a 1000 µl pipette into a 1.7 ml tube. The tube was spun and resultant supernatant was discarded.

2.3.15 DNA Library Bead Enrichment

2.3.15.1 Preparation For Enrichment

The heated dry-block temperature was set at 65 ºC. The Melt Solution was prepared by mixing 125 µl of NaOH (10N) in 9.875 ml of molecular biology grade water. 1 ml of Melt solution was added to the 1.7 ml tube of beads and was vortexed. The contents were left at room temperature for 2 mins. 1 ml of Annealing buffer was added to the 1.7ml tube of beads. The tube was centrifuged and supernatant was discarded. 45 µl of annealing buffer, 15 µl of Enrich primer A and 15 µl of Enrich Primer B were added to the 1.7 ml tube of beads. The tube was placed in a heating block at 65 ºC for 5 min and was cooled on ice for 2 mins immediately. 1 ml of Enhancing buffer was added to the 1.7 ml tube of beads and were vortexed, centrifuged and supernatant was discarded and repeat this nine more times. The tube was left at room temperature for a few minutes.

2.3.15.2 Preparation Of The Enrichment Beads

The tube of brown enrichment beads was vortexed for 1 min to resuspend the contents. The tube was placed in a Magnetic Particle Concentrator (MPC) and left for 3 min for the enriched beads to pellet. The supernatant was discarded. 500 µl of Enhancing buffer was

added and were vortexed. The previous step was repeated and finally 80 µl of Enhancing buffer was added and vortexed.

2.3.15.3 Enrichment Of The DNA-Carrying Beads

80 µl of washed enrichment beads were added into 1.7ml of tube of beads and were vortexed. The tube was rotated on LabQuake at room temperature for 5 min. The tube was placed on MPC and left for 5 mins to pellet the Enrichment beads. The supernatant was discarded carefully. The beads were washed with enhancing buffer several times until no visible beads were left in the supernatant.

2.3.15.4 Collection Of The Enriched DNA Beads

The enriched bead pellets were resuspended in 700 µl of Melt solution. The tube was vortexed for 5 sec and was left on MPC to pellet the beads. This step was repeated twice and finally 1 ml of annealing buffer was added and vortexed for 5 sec. 100 µl of annealing buffer was added finally.

2.3.15.5 Seq Primer Annealing

15 µl of Seq Primer A and 15 µl of Seq Primer B was added and vortexed. The tubes were placed in a heat block at 65 ºC for 5 min and were cooled on ice for 2 min. 1 ml of annealing buffer was added and vortexed for 5 sec. The tube was centrifuged and supernatant was discarded. This step was repeated thrice. GS Junior sequencing run required 500,000 enriched beads. The amount of enriched beads was calculated using GS Junior Bead Counter. The beads were stored at 2 to 8 ºC and were sequenced.

2.3.15.6 The GS Junior Bead Counter

The requirement for GS Junior is 500,000 enriched beads. To count the number of enriched beads, GS junior counter was used.

2.4 Denovo Sequence Assembly

The resulting DNA sequences from both Illumina and Roche 454 were processed using CLC Genomic workbench version 5.1 software to demultiplex the sequences into sets containing all sequences with the same index tag and to remove both the spacer (A) and the index tag of 6 nucleotides from the final sequences. The redundancy filter was applied by CLC version 5.1. The processed sequences, which have odd sequences, were removed using trim settings based on quality scores, ambiguity (stretches of Ns was removed) and length of reads (reads which were shorter or longer than a specified threshold were removed). Finally, the reads which were less than 500 nucleotides long were discarded. The trimming of N's from end, sequences with internal N's and trimming reads on quality was applied.

2.4.1 Denovo Sequence Analysis And RAST Annotation

The final sets of sequences for each tag were subjected to de novo sequence analysis resulting in the format of a number of contiguous (contigs). The denovo assembly algorithms used de Brujin graph. This is a type of graph which made overlap/layout/consensus graph. The contig sequences were built using the information in the read sequences. All reads were mapped using the simple contig sequence as reference and is shown in Figure 2.11 and 2.12. A genome is comprised of long stretches of contigs DNA sequences (base pairs). The contigs were exported from CLC as FASTA files and these were annotated using the RAST system which enables comparison of isolates on a gene-by-gene basis. For comparative purposes, the MG1 sequence as a reference was

annotated using RAST (http://rast.nmpdr.org/) with the previously identified gene calls retained but with new RAST annotation release version 59. In effect the genes identified by the CMR website http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi were retained but annotated using the methodology employed in RAST. K20, c505, OT170, OT171 and OT175 were also obtained from the publically available database (www.homd.org) and included in the study. Figure 2.13 represents the flow chart describing the data analysis steps used in NGS sequencing. From this the numbers of contigs per genome was obtained and then were all annotated using the same annotation software so facilitating the comparison of genomes. Preliminary gene-by-gene analysis was carried out using RAST.

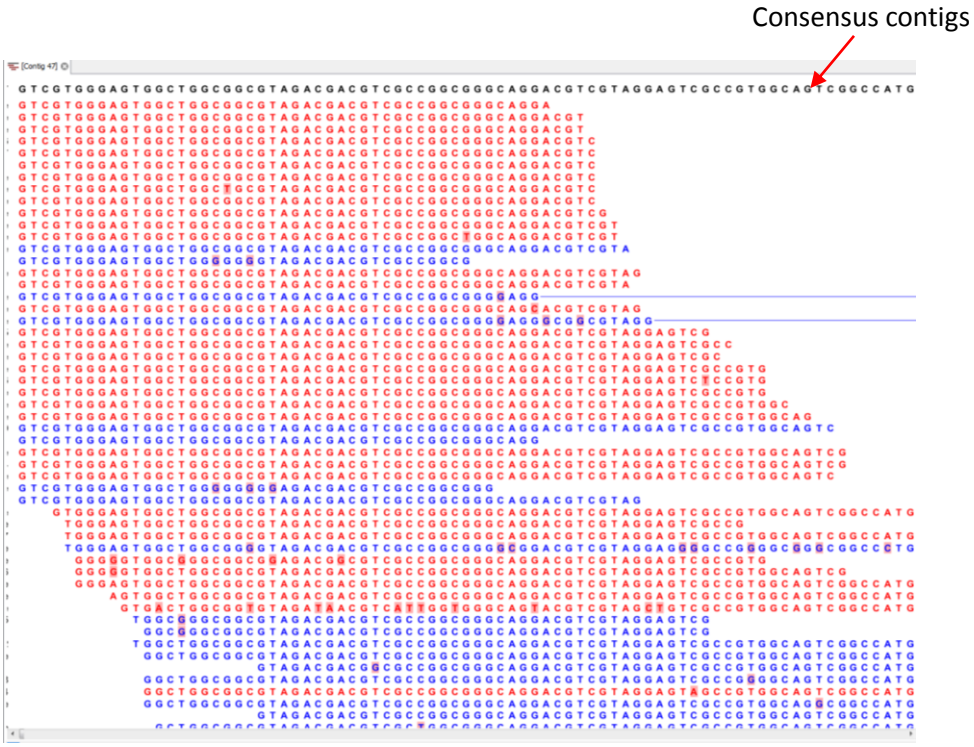


Figure 2.11: The base level view of the assembled paired-end reads (contigs)

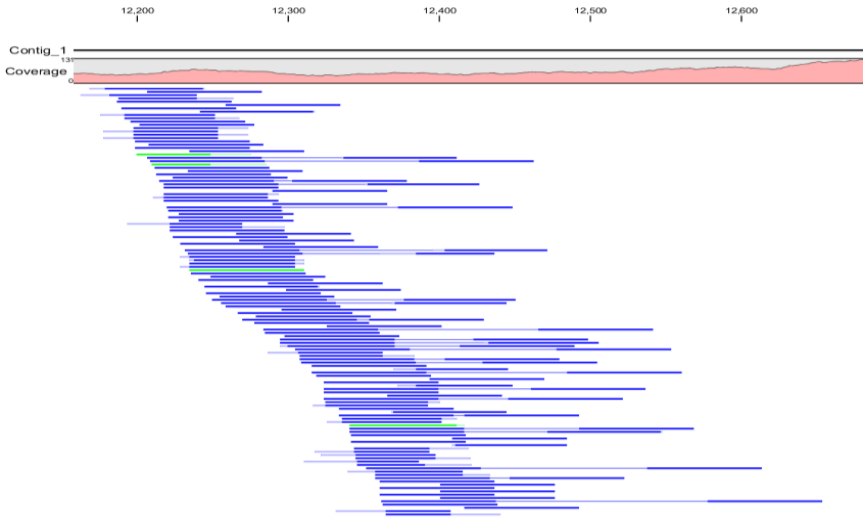


Figure 2.12: A distant view of the assembled paired-end reads to form a contig

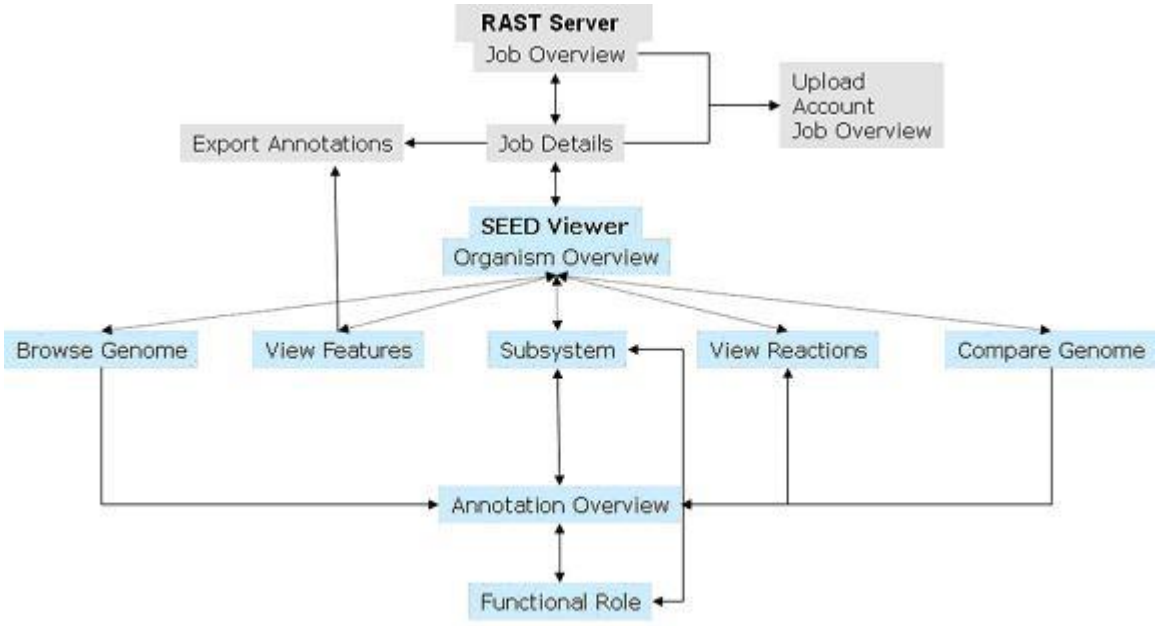


Figure 2.13: The aligning software and implication flowchart to align genome sequences

***(**http://www.theseed.org/wiki/SEED_Viewer_Tutorial**)**

2.4.2 Example of Annotation Using RAST

2.4.2.1 Comparison Of Genes of *A. naeslundii* And *A. oris*

After the denovo assembly of strains, the contig sequences obtained were exported in FASTA format from CLC and annotated in RAST (Rapid Annotation Subsystem Technology). The annotated genomes were inspected in SEED viewer enabling the comparison of strains with reference strain of MG1 (http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi), which is an *A. oris* strain. The urease region in *A. oris* and *A. naeslundii* strains were compared and it was seen that MG1 (*A. oris*) lacked the whole urease operon while c505 and S64C (*A. oris*) and R13240 and MMRC12 (*A. naeslundii*) contained urease genes in its genome (Figure 2.14).

1			2			3			4			5			
10	59	291	bi	1	184	bi	97	790	bi	9	2271	bi	21	194	Formyltetrahydrofolate deformylase
10	60	277	-			bi	97	789	bi	9	2272	bi	21	195	Urea transporter
10	61	272	-			bi	19	169	bi	9	2273	bi	21	196	UreD
10	62	227	-			bi	19	170	bi	9	2274	bi	21	197	UreG
10	63	248	-			bi	19	171	bi	9	2275	bi	21	198	UreF
10	64	164	-			bi	19	172	bi	9	2276	bi	21	199	UreE
10	65	571	-			bi	19	173	bi	9	2277	bi	21	200	Urease alpha subunit
10	66	104	-			bi	19	174	bi	9	2278	bi	21	201	Urease beta subunit
10	67	105	-			bi	19	175	bi	9	2279	bi	21	202	Urease gamma subunit
10	68	41	bi	1	185	bi	19	176	bi	9	2280	bi	21	203	Lp36p
10	69	84	bi	1	186	bi	19	177	bi	9	2281	bi	21	204	Lp31p
10	70	102	bi	1	187	bi	19	178	bi	9	2282	bi	21	205	S14p
10	71	59	bi	1	188	bi	19	179	bi	9	2283	bi	21	206	Lp33p
10	72	83	bi	1	189	bi	19	180	bi	9	2284	bi	21	207	Pp28p
10	73	402	bi	1	190	bi	19	181	bi	9	2285	bi	21	208	Hypothetical protein
10	74	437	bi	1	191	bi	19	182	bi	9	2286	bi	19	164	Serine hydroxymethyltransferase
10	75	299	bi	1	192	bi	19	183	bi	9	2287	bi	19	165	Methylenetetrahydrofolate
10	76	572	bi	1	193	bi	19	184	bi	9	2288	bi	19	166	

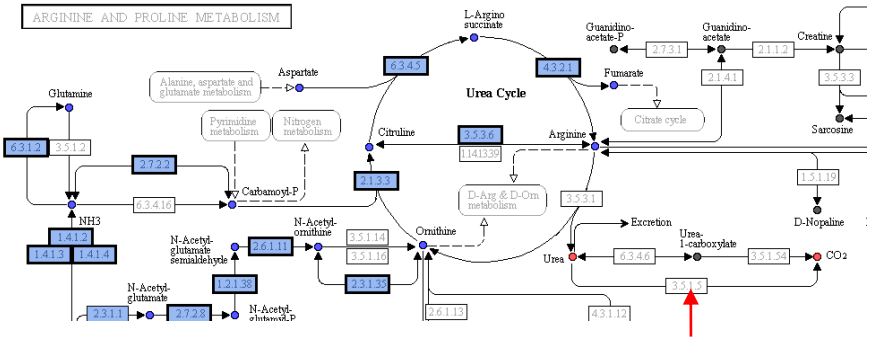
Figure 2.14: Comparison of urease gene region in 5 isolates of *A. oris* and *A. naeslundii*

***(1- *A. naeslundii* [R13240]; 2-*A. oris* [MG_1]; 3-*A. naeslundii* [MMRC12-1]; 4-*A. oris* [c505]; 5-*A. oris* [S64C]**

2.4.2.2 Metabolic Reconstruction Using KEGG

The metabolic pathway of strain *A. oris*- MG1 (*A. oris*) and R13240 (*A. naeslundii*) was reconstructed using KEGG software tool from within the RAST program and it showed that urease enzyme was lacking in the metabolic pathway of *A. oris*-MG1 while it was present in R13240 (Figure 2.15).

A



B

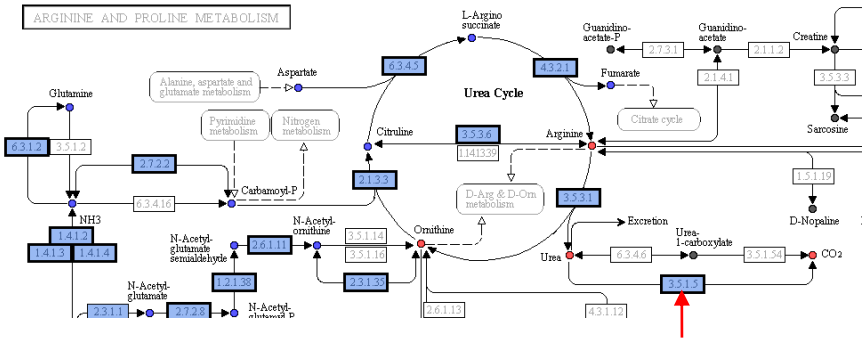


Figure 2.15: Comparison of MG_1 and R13240 by constructing metabolic pathway.

(A- *A. oris* - MG1; B - *A. naeslundii* R13240; - 3.5.1.5 urease enzyme with arrows pointing towards enzyme)

2.4.2.3 Comparison Of Whole Genomes Of *A. oris* and *A. naeslundii* Strains Relative To *A. naeslundii* Strain (R13240)

The whole genomes were compared using the sequence based comparison tool in RAST and it was seen that percent protein sequence identity of *A. naeslundii* (MMRC12-1) was higher relative to *A. naeslundii* (R13240) while the other three *A. oris* strains are more similar with each other as compared to R13240 (Figure 2.16). The same was observed when *A. oris* strains were compared with *A. naeslundii* strains. The protein sequence colour similarity code reveals that *A. oris* is different to *A. naeslundii* in genetic features and suggest that both are two genospecies of *Actinomyces*.

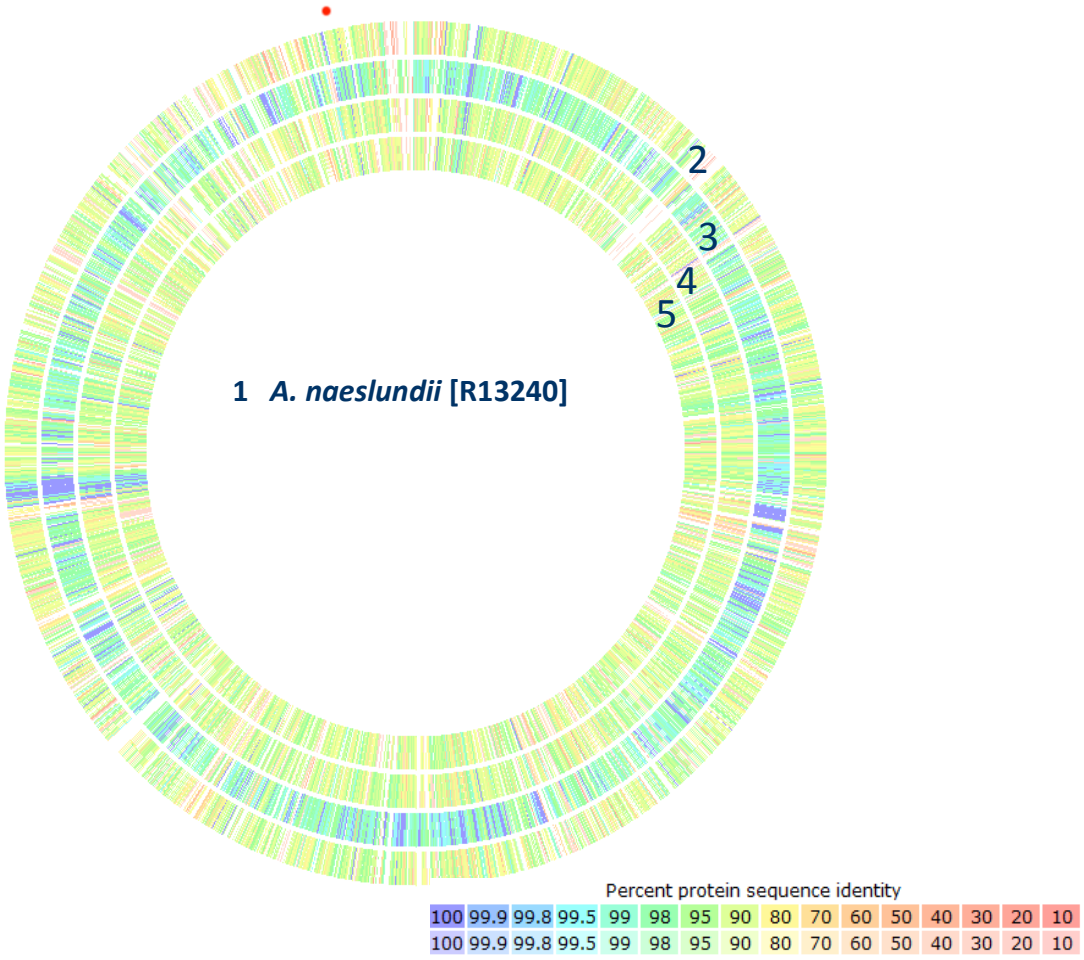


Figure 2.16: Whole genome comparison of *A. oris* and *A. naeslundii* strains

***(1- *A. naeslundii* [R13240]; 2-*A. oris* [MG_1] ;3-*A. naeslundii* [MMRC12-1]; 4-*A. oris* [C505]; 5-*A. oris* [S64C]). The colour code represents the amino acid sequence identity to reference genome. The more purple or blue colour indicates higher percent protein sequence identity**

Chapter 3 GENOME SEQUENCING

3.1 High-Throughput Whole-Genome Sequencing And Mapping

3.1.1 Description Of Isolates Selection

36 *Actinomyces* isolates were selected of which 19 belong to *Actinomyces naeslundii* and 17 belong to *Actinomyces oris* group of species.

3.1.2 Identification Of Isolates

The 36 isolates used in this study were previously identified using MLST analysis (Henssge, 2009) and grouped into 6 clusters of *A. oris* and 3 clusters of *A. naeslundii* using concatenation of 7 housekeeping gene analysis and were shown with Neighbour-Joining trees in Figure 3.1 & 3.2.

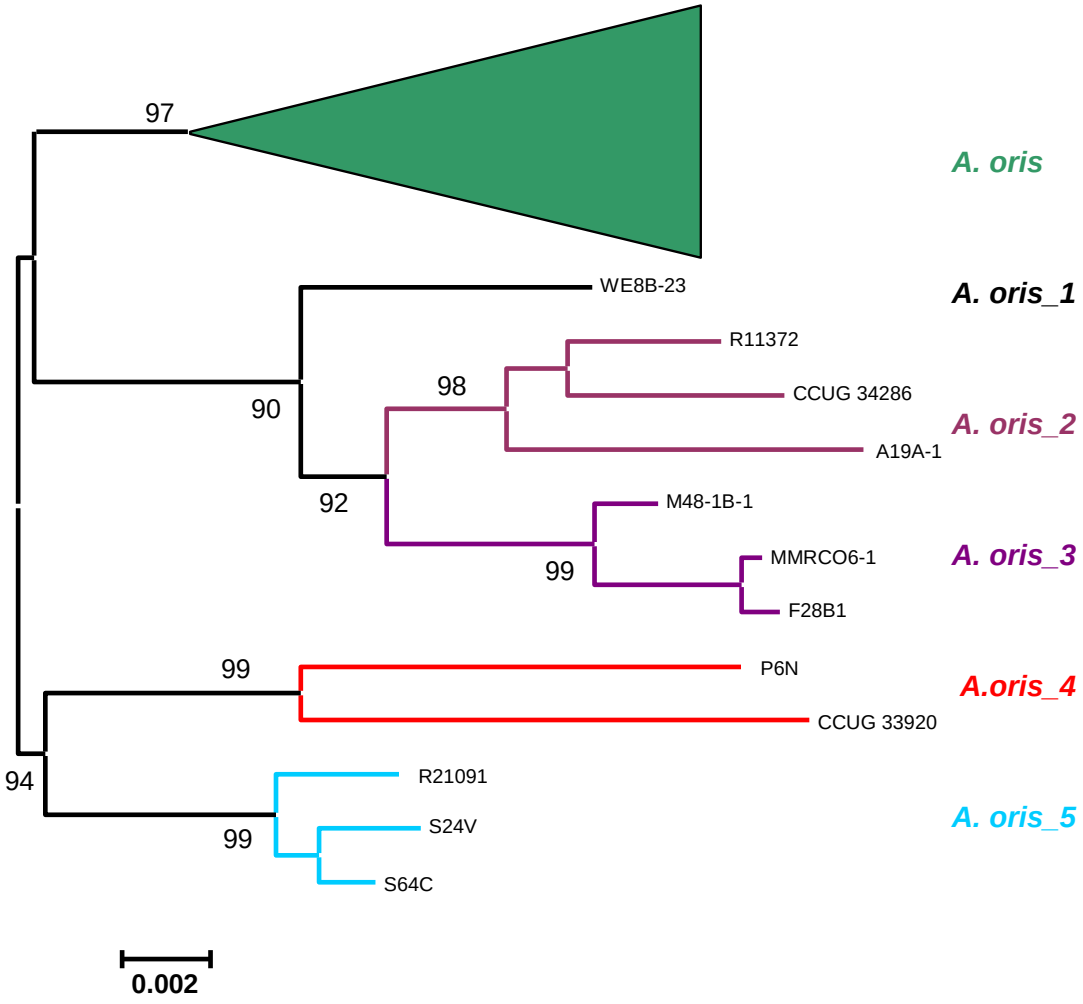


Figure 3.1: Neighbor-Joining tree of *A. oris* isolates obtained using concatenation of 7 housekeeping genes (taken with modification Henssge *et al*, 2011).

***The values are bootstrap values. These isolates further subjected to whole genome sequencing.**

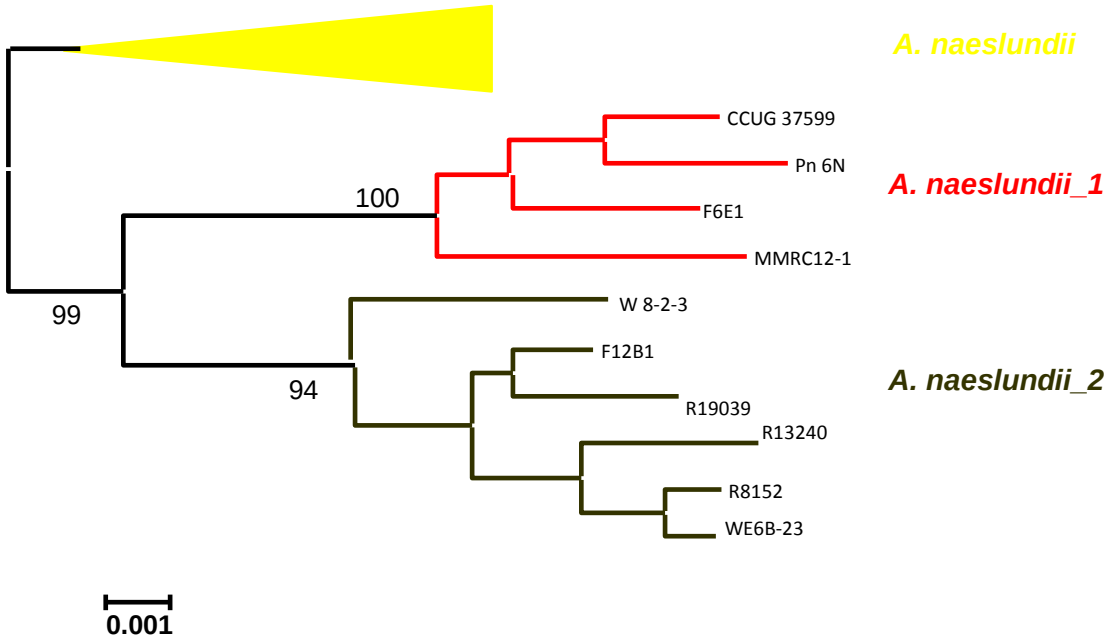


Figure 3.2: Neighbor-Joining tree of *A. naeslundii* isolates obtained using concatenation of 7 housekeeping genes (taken with modification (Henssge *et al*, 2011)).

*** The values are bootstrap values. These isolates were further subjected to whole genome sequencing.**

3.1.3 High Throughput Whole-Genome Sequencing Of Selected *Actinomyces* Strains To Dissect The Genomic Diversity Among Isolates

The 5 µg of whole genomic DNA was extracted using protocols described in chapter 2. Genomic DNA samples were run on Illumina and Roche 454 Genome Sequencers.

Fragment (76bp) and Multiplexed Paired-end (76 x 2) libraries were generated. 12 index tags were used for the 12 samples and were run in a single lane. The same indexing was repeated for the next 12 samples. Demultiplexing of tag reads was done using CLC genomics workbench version 5.1.

3.1.4 Importing qseq Files From Illumina Genome Sequencer

The raw sequences were stored on the server after the analysis was completed by the Illumina Genome Analyzer. The sequences were imported from the server using tools on CLC Genomics workbench 5.1 called NGS import and files were selected containing both forward and reverse reads e.g. "S_1_2 and "S_1_2" and saved in a new folder. They were resulted into a group of 13 (12 grouped with index tags and one was ungrouped).

3.1.5 Process Tagged Sequences

The tag sequences were processed using option "Process tagged sequences" for high throughput analysis from CLC Genomics work bench version 5.1. 120 paired reads of one lane were selected. The input information about the sequences was choosen in the following order "Sequences 1 to 76, Barcode 6, Linker 1"; and then "Barcode 6, Linker 1, Sequences 1 to 76". For Barcode 1 "CGTGAT", the reverse complement "GCACTA" was used and then the order of sequence was reversed "ATCACG". They were put twice "ATCACG-ATCACG" and saved in a new folder. Same was repeated for all 12 barcodes. The barcodes are listed in Table 3.1.

Table 3.1: The barcodes used in the processing of tagged sequences

Index Tag	Reverse Complement of index tag	Reverse order of *RCIT
CGTGAT	GCACTA	ATCACG-ATCACG
ACATCG	TGTAGC	CGATGT-CGATGT
GCCTAA	CGGATT	TTAGGC-TTAGGC
TGGTCA	ACCAGT	TGACCA-TGACCA
CACTGT	GTGACA	ACAGTG-ACAGTG
ATTGGC	TAACCG	GCCAAT-GCCAAT
GATCTG	CTAGAC	CAGATC-CAGATC
TCAAGT	AGTTCA	ACTTGA-ACTTGA
CTGATC	GACTAG	GATCAG-GATCAG
AAGCTA	TTCGAT	TAGCTT-TAGCTT
GTAGCC	CATCGG	GGCTAC-GGCTAC
TACAAG	ATGTTC	CTTGTA-CTTGTA

* RCIT (Reverse Complement of index tag)

3.1.6 Trimming Of Sequences

The redundancy filter was applied by CLC version 5.1. The processed sequences were trimmed based on quality of reads. The reads which were less than 500 nucleotides long were discarded. The trimming of N's from end, sequences with internal N's and trimming reads on quality was applied. The results obtained are shown in Table 3.2. There were two types of sequences used in the current study. The sequences which were obtained using Illumina Genome Analyzer (IGA) are shown with their barcode 1-12 and sequences obtained with Roche 454 platform are represented with RL1-RL8. The read length distribution was bimodal. One was giving a mean of ~76 and other was of ~400 bp. The quality of sequence reads were high but were dropped after 400bp. Sequence reads were GC rich. The first 9 bases of almost all sequences were ATATCGCGA. This was represented as an adapter sequence at the 5′ end and was trimmed.

3.1.7 Quality Trimming

Due to the higher computational resources required by CLC version 5.1, efforts were made to optimize the parameters to get the high quality results. The low quality bases present in a sequence read were misleading in the downstream applications. The base-calling of each nucleotide was considered and original sequence was trimmed in order to get longest sequence with the high quality. The N's (indetermination) were extracted out of the sequence to improve the quality.

The highest number of input reads observed were 37120172 bp for G127B (*A. oris*) using IGA while lowest number of input reads was found to contain 1392262 bp for the isolate CCUG 35334. The average number of input reads using Illumina Genome Analyzer system resulted were 12660382 ± 11238640 nucleotides, of which, sequences ranged from 0 – 819539 nucleotides were removed based on quality trimming, among which 245570 ± 577004 reads were unambiguous (do not have any undetermined nucleotide) and finally sequence had number of reads ranged between 1.37-3.7 Mbp, but once pre-

processed, the average insert size was 73.8 bp for illumina and 356 bp for Roche 454 sequences (Table 3.2).

The number of input reads for Roche 454 sequences recorded with the initial average length of reads of 67398 nucleotides. The smallest read length observed was 206 and longest read length gave 131830 nucleotides. The initial average insert size was 359.66 bp ± 9.75bp. Three trim settings were used. The average trim on quality gave 14032 ± 6508 bp; the ambiguity trim resulted in 6513 ± 9628 nucleotides and finally average length after trim was 353 ± 9.93. Finally number of reads after trim was 67398 ± 31053. The number of reads ranged between 206 -131828. Both kinds of data clearly showed that trimming of sequences ≤ 500 nucleotide gave quality reads (Table 3.2).

Table 3.2: Summary of Trimming of High Throughput Genomic Sequences

Organism	Sequencing Method	Number of input reads	Initial Average length	Sequences Trimed with low quality	Sequences Trimmed with N's (Ambiguity trim)	Avg.length after trim	Number of reads after trim
A7A-1	Illumina	25071572	76	0	108265	75.9	25071572
A7A	Roche 454	88574	359.9	18126	5430	354.3	88573
A19A-1	Illumina	3092600	76	638518	13715	71.8	3055676
A19A-1	Roche 454	64247	354.1	13115	5291	347	64247
CCUG 33920	Illumina	2818720	76	595082	12466	71.9	2800861
CCUG 33920	Illumina	9295296	76	0	34718	75.9	9295296
CCUG 33920	Roche 454	119046	370	24538	9730	362.9	119044
CCUG 34286	Illumina	25020220	76	0	112737	75.9	25020220
CCUG 34286	Roche 454	57437	364.2	12650	3578	358.6	57437
CCUG 35334	Illumina	1392262	76	252537	6202	72.3	1371459
CCUG 35334	Illumina	8567178	76	0	32461	75.9	8567178
CCUG 35334	Roche 454	49739	358.2	9839	3962	351.4	49739
CCUG 37599	Illumiina	1467882	76	278132	6693	71.8	1439189
CCUG 37599	Illumina	8434366	76	0	31610	75.9	8434366
CCUG 37599	Roch 454	131830	362.1	25838	10099	355.5	131828
F4D1	Illumina	26638378	76	0	1727548	74.6	26638378
F4D1	Roche 454	56967	364.8	12792	3962	358.5	56966
F6E1	Illumina	36471366	76	0	156002	75.9	36471366
F6E1	Roche 454	106374	366.9	23663	6556	361.5	106374
F12B1	Illumina	3639846	76	629960	16294	72.4	3597478
F12B1	Roche 454	83366	359.6	16568	6569	353	83365
F12B1	Illumina	8826486	76	0	33417	75.9	8826486
F28B1	Illumina	3306080	76	714137	14440	71.6	3269374
F28B1	Roche 454	68861	356.1	13991	63143	349.2	68861
G53E	Illumina	3744952	76	819539	16695	71.5	3700196
G53E	Roche 454	43306	333.2	8522	3326	326.7	43306

G127B	Illumina	37120172	76	0	2376149	74.5	37120172
G127B	Roche 454	49306	366.4	10875	3196	360.6	49306
M48-1B-1	Illumina	3028280	76	659439	13614	71.6	2994860
M48-1B-1	Roche 454	80809	343	15786	6367	336.4	80805
MB-1	Illumina	26819316	76	0	113243	75.9	26819316
MB-1	Roche 454	24452	364	5212	1800	357.4	24452
MMRC12-1	Illumina	3062774	76	635575	13483	71.7	3011740
MMRC12-1	Roche 454	49758	347.5	9637	3891	341.2	49757
MMRCO6-1	Illumina	26652946	76	0	118689	75.9	26652946
MMRCO6-1	Roche 454	70499	368.8	15615	4624	70499	70499
NCTC 10301	Illumina	3044374	76	537945	13704	72.5	3015053
NCTC 10301	Illumina	11382784	76	0	42421	75.9	11382784
NCTC 10301	Roche 454	43610	364.2	8849	3576	357.2	43610
P6N	Illumina	2276512	76	411442	10507	72.4	2252461
P6N	Illumina	7289862	76	0	27750	75.9	7289862
P6N	Roche 454	45858	359	9634	3934	351.7	45858
Pn6N	Illumina	13321842	76	0	57911	75.9	13321842
Pn6N	Roche 454	33361	358.7	6784	2036	353.4	33360
R8152	Illumina	12266662	76	0	55638	75.9	12266662
R8152	Roche 454	84289	360.5	17590	5061	355.2	84288
R11372	Illumina	3154048	76	668891	13978	71.5	3112152
R11372	Roche 454	117792	358.3	24635	9708	351.4	117792
R13240	Illumina	3161894	76	696191	13882	71.7	3129717
R13240	Illumina	4661492	76	0	17651	75.9	4661492
R13240	Roche 454	96145	360.9	18905	7473	354.3	96145
R19039	Illumina	27503178	76	0	118817	75.9	27503178
R19039	Roche 454	54221	364.9	11676	3331	359.6	54220
R21091	Illumina	9349216	76	1755715	42583	72.3	9240370
R21091	Illumina	8539910	76	0	32083	75.9	8539910
R21091	Roche 454	66094	349	13315	5253	342.3	66093

R23275	Roche 454	43236	364.5	9656	2997	358.4	43235
R23275	Roche 454	206	321.3	25	11	314.3	206
R23275	Illumina	10309720	76	0	683144	74.5	10309720
R23275	Illumina	31172444	76	0	134085	75.9	31172444
R24330	Illumina	34279406	76	0	144114	75.9	34279406
R24330	Roche 454	25350	360.9	5233	1677	355	25350
S24V	Illumina	30725374	76	0	1974285	74.6	30725374
S24V	Roche 454	114665	363	26901	7817	356.7	114665
S43L	Illumina	32511482	76	0	2078402	74.5	32511482
S43L	Roche 454	42023	366.1	8938	2750	360.3	42023
S44D	Illumina	12683538	76	0	54638	75.9	12683538
S44D	Roche 454	53993	367	11585	3371	361.5	53993
S64C	Illumina	2630122	76	596916	11698	71.5	2608409
S64C	Roche 454	83487	363.5	16371	6613	356.7	83487
S65A	Illumina	2991652	76	618091	13291	71.8	2955464
S65A	Roche 454	49655	365	9967	3920	358.3	49655
T23P	Illumina	22718302	76	0	95923	75.9	22718302
T23P-1	Roche 454	51014	362	10847	3472	356	51014
W8-2-3	Illumina	3262748	76	679555	14334	71.7	3216127
W8-2-3	Roche 454	96674	364	19732	7621	357.3	96674
W11-1-1	Illumina	3314442	76	768020	14499	71.1	3253839
W11-1-1	Roche 454	128093	367.5	27052	10518	360.4	128093
WE6B-3	Illumina	4260272	76	751821	19467	72.5	4216354
WE6B-3	Illumina	11469454	76	0	43076	75.9	11469454
WE6B-23	Roche 454	81430	368.4	16899	6415	361.7	81429
WE8B-23	Illumina	12592162	76	0	818983	74.6	12592162
WE8B-23	Illumina	9694380	76	0	36502	75.9	9694380
WE8B-23	Roche 454	31454	362.6	6983	2336	356	31454
WE8B-23	Roche 454	73928	357.2	14893	6088	350	73928

Table 3.3: Summary of Denovo Assembly analysis using High-throughput genomic analysis

Name of Organisms	Original Contig Length	Original No. of Contigs	Maximum Contig Length	*Final N50	Total of Percent of Reads Matched	Final Coverage of Contigs	*Final Contig Length	*Final No. of Contigs	G+C content
A19A-1 (AO)**	3301420	126	167536	54297	98	65	3283870	118	67.9
W11-1-1 (AO)	2998878	54	228602	156506	93	86.4	2976384	45	68.8
G53E (AO)	3044774	41	369220	202579	92	81.8	3040072	37	68.7
S64C (AO)	3127055	103	210279	72461	95	65.4	3114492	96	68.5
F28B1 (AO)	3294380	113	193841	55827	94	75.6	3278855	103	68
M48-1B-1 (AO)	3256245	101	203715	82914	94	69.58	3243259	90	68
R11372 (AO)	3295754	84	229462	81752	93	71.9	3282203	74	67.9
CCUG 33920 (AO)	2972582	46	342576	116999	96	308	2972582	45	68.2
WE8B-23 (AO)	3520872	407	194071	95548	99	294	3229736	88	67
R21091 (AO)	3071193	106	196712	76730	94	415	3046442	87	68.7
P6N (AO)	3120358	49	251637	116121	95	224	3117455	46	68.1
S24V (AO)	3091335	309	82497	22701	95	12.3	3045527	291	68
F4D1 (AO)	3070322	87	283407	110490	98	615	3026188	66	67.5
R23275 (AO)	3292170	208	186811	59254	98	907	3115162	96	68.3
MMRCO6-1 (AO)	3296052	99	207902	74046	98	594	3250825	86	68
CCUG 34286 (AO)	3320662	167	333772	108205	98	626	3275811	91	67.9
A7A-1 (AO)	3157501	116	257632	151307	98	644	3009476	42	68.6
G127B (AN)***	3153606	54	354722	140660	98	848	3139241	42	67.9
MB-1 (AN)	3242122	97	321450	192595	98	608	3201924	42	67.8
R24330 (AN)	3163414	88	176599	87720	98	791	3139836	61	67.9
T23P-1 (AN)	3129692	129	204106	40421	98	510	3068220	104	67.9
S44D (AN)	3134327	61	232455	128603	98	313	3128129	52	67.9
F6E1 (AN)	3093257	191	121099	35559	96	11.1	3041208	174	67.5
R19039 (AN)	3275431	171	403515	102825	98	650	3197928	56	67.4
R8152 (AN)	3202166	60	264847	143743	98	311	3183492	44	67.5

Pn6N (AN)	3163543	63	481508	109285	98	305	3155768	53	67.5
W8-2-3 (AN)	3229018	52	386802	95588	93	103	3220041	50	67.6
S65A (AN)	3237636	37	240210	160271	93	63	3229861	33	66.8
MMRC12-1 (AN)	3177225	63	325091	117741	92	65	3165113	56	67.6
R13240 (AN)	3228270	33	393691	258514	96	203	3227505	32	67.6
CCUG 35334 (AN)	3162717	54	324971	136770	96	236	3151111	47	67.7
CCUG 37599 (AN)	3165585	58	273969	87807	96	236	3159810	57	67.5
F12B1 (AN)	3198051	81	251509	91878	95	306	3192008	73	67.6
WE6B-3 (AN)	3234737	63	333572	106225	92	340	3222459	58	67.6
NCTC 10301 (AN)	3128554	52	271023	149929	93	339	3119690	47	67.9
S43L (AN)	3296594	295	324916	220568	98	732	3122805	34	67.8

***** Final indicates sequences after removal of short reads and reads with low coverage and reads not matched by Blast to bacteria
** AO: *A. oris*
*** AN: *A. naeslundii*

3.2 The Assembly Approaches For The *Actinomyces oris* Genome

3.2.1 Denovo Assembly Of Short And Long (combined) Sequence Reads

The genomes of 36 *Actinomyces* strains were subjected to denovo sequence analysis and results were shown in Table 3.3. The total length of the sequences per genome obtained was between 2.9 – 3.5 Mbp. The number of contigs was low in all sequenced strains. The minimum number of contigs was found to present in isolate name R13240 (AN), which had 32 numbers of contigs. The maximum number of contigs was 295 in isolate name S43L (AN). All other isolates have number of contigs ranging between 32 and 291. The maximum contig length was 481508 nucleotides in isolate Pn6N (AN). The minimum contig length was 82497 nucleotides in isolate S24V (AO). The N50 was calculated to capture how much of the assembly is covered by relatively large contigs. To compute N50, the list of length was first ordered from low to high. Then the sequence was summed from the longest sequence until reached to a one-half of the total length of all the contig sequence. This sequence gained 50% weight of all the contigs length therefore the corresponding contig length was N50. The N50 of trimmed sequences was in the range of 22701 - 258514. The proportion of reads matched was also calculated by dividing total read length with the matched read length. The proportion found was between 92 to 98%. The average coverage of each contig was obtained after denovo assembly analysis for each isolate and was shown in Table 3.3. Final count after removing the sequences from denovo analysis was in the range of 32 - 291. Finally, the length of genome after denovo assembly was in the range of 2.9 - 3.2 Mbp. The average G+C content was 67-68 %. T- test was used to compare the (G+C) content of AO and AN. The mean (G+C) content for the AO was 68.12 (± 0.46) which is significantly greater than that for AN with (G+C) of 67.63 (± 0.26) ($t = 3.878$; $p = 0.0006$). The t test was also used to compare the significant difference in the mean final contig length and it was found that the mean of final contig length of AO was

3135785 (± 118207) which is not significantly different to that of AN with contig length of 3161376 (± 52281) (p=0.419;t=0.847).

3.2.2 Identification Of Un-assembled Sequences After Denovo Assembly

The sequences found with greater cover than the remainder of the genome were identified as transposases, 16S or 23S genes. These were removed. The reason to remove these sequences and not include in contigs was because they were clearly present as multiple copies in the genome and from the data available it was unable to determine their location.

3.2.3 Number Of Genes Versus Size of Genome

The numbers of genes present in a genome were plotted against the genome size of denovo assembled *Actinomyces* strains and the results are shown in Figure 3.3. T-test was applied. The results clearly indicate that maximum number of genes were present in a genome of maximum length and has a significant correlation with correlation coefficient of 0.91. The combined consensus contig length increases therefore suggests that RAST was able to identify more coding regions. This indicates that assembly is meaningful and not simply artifactual.

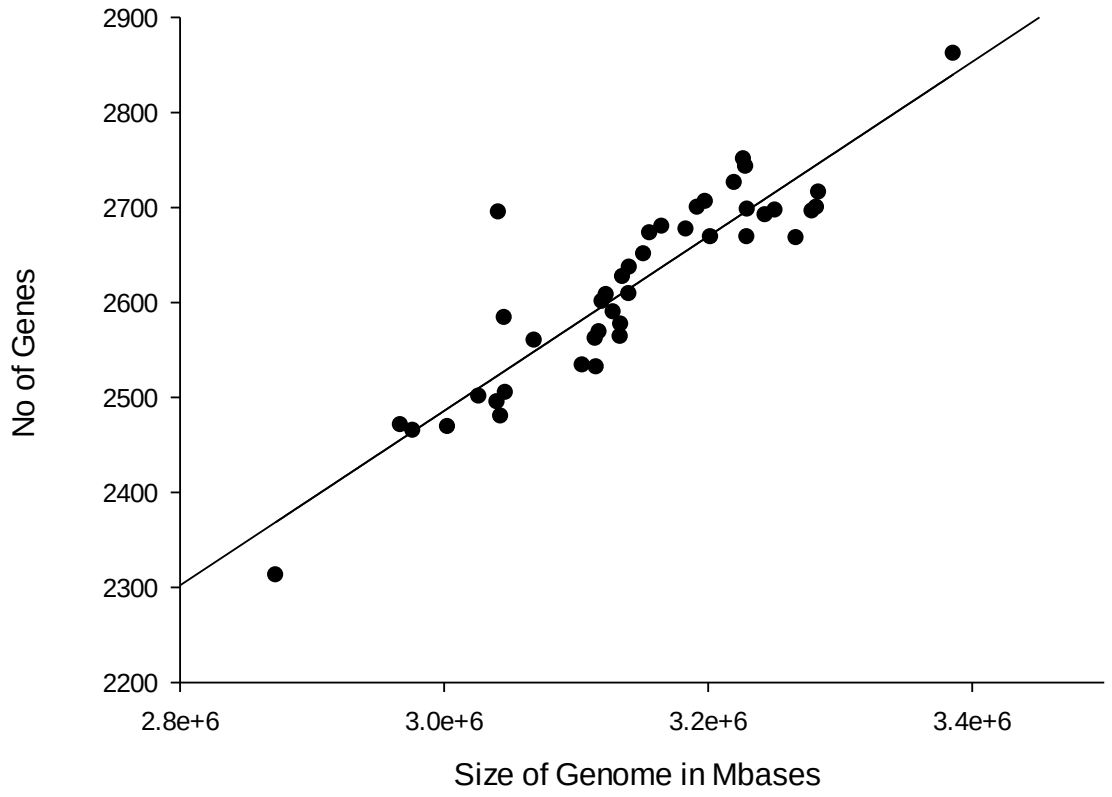


Figure 3.3: Genome size (Mbases) plotted versus number of genes RAST

3.3 Map-to-Reference (MG1) Assembly Of Highthroughput Short And Long Reads (Combined) Sequences

The 36 *Actinomyces oris* and *Actinomyces naeslundii* strains (Table 2.1) data set is made up of 2.3-2.8 million 76-base pair reads by Illumina and 400-base pairs by Roche 454. The sequences were combined and trimmed. Since *Actinomyces oris*-MG1 strain was already sequenced and assembled and available online on HOMD website, the published sequence was used as a reference to evaluate the accuracies of the assemblies prepared in this study. The reference *A. oris*-MG1 genomic sequence length measured was 3.04 Mbp and has only one contig (complete genome is available) (NCBI Taxon ID 240017). Consensus sequence of *A. oris* -MG1 was

extracted using CLC genomic workbench tool and compared with sequenced strains of *Actinomyces*.

3.4 Annotation Using RAST

3.4.1 Annotation Of Denovo Assembled Strains Of *Actinomyces*

The genes on each contig were annotated using the Rapid annotation subsystem technology version 4.0 (http://rast.nmpdr.org), and Seed viewer programme. To be considered valid the denovo assembly was performed and minimum contigs of 500 bases were kept. The output file is a list of contig sequences. The file containing contigs (.faa) was extracted out of CLC. These (.faa) files were submitted to RAST server for annotations. The 36 sequence (.faa) files were uploaded using software RAST version 4.0. The version 11 of the genetic code was selected for bacteria. Genus, species and strains names were given. RAST 59 release was used to compute the sequences. The organism's details were viewed under "Browse annotated genomes in SEED viewer". The size of genome was found to be within the range of 2.9 to 3.2 Mbps. The numbers of contigs observed were in the range of 32-291. The numbers of subsystem found were in the range of 251-305. The numbers of coding sequences present were in the range of 2313 - 2751. The numbers of RNA's were found to be in the range of 46-60.

The seven sequenced files from the internet database (NCBI) were uploaded of the following micro-organism OT171, OT170, OT175, MG1, k20, c505 and *A. johnsonii*. These were annotated using RAST using the same settings as described above. They have genome size in the range of 2.8-3.1 Mbs and contig counts was in the range of 1-771. The summary of the output from RAST annotation is given in Table 3.4.

3.4.2 Annotation (RAST) Output Of Map-to-Reference Assembled Strains

The RAST output of Map to Reference assembled strains was also taken (Table 3.5). The smallest genome size found was of MB-1 (AN) which is 2.3 Mbases and the length of all AN genome were in the range of 2.3-2.4 Mb while all AO strains had greater genome size as compared to AN isolates genome and in the range of 2.5 – 2.8 Mbases. *A. oris* isolates genomes had a higher number of subsystems ranging between 367 - 386 while *A. naeslundii* subsystems were in the range of 351-366. The number of coding sequences found was in the range of 2176-2414. Only 7 genes were found to be missing in isolate R8152 (AN) while 30 genes were missing in CCUG 34286 (AO) while annotating isolates using RAST (Table 3.5). Alignment of this assembly to the *A. oris*-MG1 reference genome indicated that 98% of the genome was covered with identity of 99 %, and few misassemblies (i.e., juxtaposition in scaffolds of non-contiguous regions of the genome) were observed, indicating a high quality draft genome for *Actinomyces.*

3.4.3 Comparison of Genes From Denovo Assembled Strains Versus Map-to-Reference Assembled Strains

The comparison of coding sequences obtained after annotation of denovo assembled strains with the Map-to-reference assembled strains were carried out. The number of coding sequences obtained were between 2176-2414 in the Map-to-reference assembled strains and 2313-2862 coding sequences were obtained with denovo assembled strains. This showed that denovo assembly approach is better as compared to Map-to-reference assembly becasue it gave more genes and showed high quality draft genome.

Table 3.4: RAST Output after annotation of denovo assembled selected strains

Name of Organisms	Size of Genome	No. of Contigs	No.of subsystems	No.of coding sequences	Possibly missing genes	No. of RNA's
A7A-1 (AO)	3104690	42	296	2534	44	53
A19A-1 (AO)	3283870	118	300	2716	65	47
CCUG 33920 (AO)	2966880	45	287	2471	34	51
CCUG 34286 (AO)	3266848	91	297	2668	57	49
WE8B-23 (AO)	3229515	88	299	2669	67	52
F4D1 (AO)	3026188	66	297	2501	68	53
F28B1 (AO)	3278855	103	299	2696	95	52
G53E (AO)	3040072	37	296	2495	44	52
M48-1B-1 (AO)	3243259	90	298	2692	59	60
MMRCO6-1 (AO)	3250825	86	291	2697	59	49
P6N (AO)	3117455	46	293	2569	70	56
R11372 (AO)	3282203	74	294	2700	91	53
R21091 (AO)	3046442	87	291	2505	68	51
R23275 (AO)	3115162	96	289	2532	71	47
S24V (AO)	3045527	291	266	2584	96	54
S64C (AO)	3114492	96	285	2562	82	55
W11-1-1 (AO)	2976384	45	295	2465	45	50
CCUG 35334 (AN)	3151111	47	300	2651	64	46
Pn6N (AN)	3155768	53	300	2673	50	53
R8152 (AN)	3183492	44	304	2677	57	54
G127B (AN)	3140302	42	304	2637	64	48
R13240 (AN)	3226920	32	301	2751	54	56
R19039 (AN)	3197928	56	303	2706	74	53
MMRC12-1 (AN)	3165113	56	305	2680	54	49
F6E1 (AN)	3041208	174	290	2695	64	51
R24330 (AN)	3139836	61	299	2609	68	53
MB-1 (AN)	3201924	42	304	2669	62	55
S44D (AN)	3128129	52	300	2590	71	54
NCTC 10301 (AN)	3119690	47	305	2601	68	50
S65A (AN)	3229861	33	300	2698	74	49
T23P-1 (AN)	3068220	104	294	2560	47	47
W8-2-3 (AN)	3220041	50	303	2726	73	52
F12B1 (AN)	3192008	73	304	2700	83	54
WE6B-3 (AN)	3228491	58	304	2743	67	52
S43L (AN)	3122805	34	302	2608	59	52
MG1 (AO)	3042856	1	292	2480	35	60
c505	3133750	111	294	2577	92	60
K20	2872429	771	251	2313	123	53
OT 171	3002669	580	261	2469	112	53
OT175	3133330	7	301	2564	56	60
OT170	3135160	99	303	2627	65	51
A. johnsonii	3385844	255	284	2862	85	60
CCUG 37599 (AN)	3159810	57	299	2690	64	52

Table 3.5: RAST output after annotations of selected Map-to-reference (MG1) strains of *Actinomyces*

Name of Organisms	Size of Genome	No. of Contigs	No. of subsystems	No. of coding sequences	Possibly Missing genes	No. of RNA's
A7A-1 (AO)	2707775	1	379	2297	12	60
A19A-1 (AO)	2693904	1	373	2330	24	60
CCUG 33920 (AO)	2606045	1	375	2298	13	60
CCUG 34286 (AO)	2612487	1	367	2213	30	60
WE8B-23 (AO)	2848542	1	386	2412	19	60
F4D1 (AO)	2703424	1	378	2289	17	60
F28B1 (AO)	2451932	1	363	2269	10	60
G53E (AO)	2759021	1	383	2375	23	60
M48-1B-1 (AO)	2660554	1	376	2303	10	60
MMRCO6-1 (AO)	2619925	1	373	2229	11	60
P6N (AO)	2628134	1	371	2386	14	60
R11372 (AO)	2649235	1	378	2291	21	60
R21091 (AO)	2610430	1	376	2362	23	60
R23275 (AO)	2753926	1	376	2350	20	60
S24V (AO)	2528469	1	368	2204	15	60
S64C (AO)	2552565	1	360	2208	17	60
W11-1-1 (AO)	2760502	1	379	2414	24	60
CCUG 35334 (AN)	2438477	1	354	2288	12	60
Pn6N (AN)	2398240	1	369	2273	16	60
R8152 (AN)	2406774	1	367	2186	7	60
G127B (AN)	2450266	1	363	2277	7	60
R13240 (AN)	2475034	1	361	2252	9	60
R19039 (AN)	2397884	1	364	2258	13	60
MMRC12-1 (AN)	2482166	1	365	2348	14	60
F6E1 (AN)	2409678	1	366	2183	7	60
R24330 (AN)	2389853	1	355	2332	4	60
MB-1 (AN)	2358500	1	358	2276	14	60
S44D (AN)	2388759	1	351	2176	12	60
NCTC 10301 AN	2453646	1	363	2326	8	60
S65A (AN)	2453041	1	365	2238	15	60
T23P-1 (AN)	2384891	1	362	2210	5	60
W8-2-3 (AN)	2539588	1	369	2349	24	60
F12B1 (AN)	2451932	1	363	2269	10	60
WE6B-3 (AN)	2444625	1	355	2266	5	60
S43L (AN)	2470367	1	366	2317	15	60
CCUG 37599 (AN)	2434059	1	366	2185	17	60

*MG1, c505, k20, OT171, OT175, OT170 and *A. johnsonii* have the same values as was shown in Table 3.4

3.5 Comparison Of Subsystem Using RAST Services

The comparison of subsystem was done using the RAST server annotation services. The subsystem of *Actinomyces oris*-MG1 was compared to *A. naeslundii*-NCTC 10301 and come across with the genes which were present in *A. naeslundii* - NCTC 10301 but absent in *A. oris*-MG1. The genes are shown in Appendix 3A. Trehalose Uptake and Utilization subsystem was found to comprise an operon containing 6 genes and the roles of these genes were Glucose/mannose: H+ symporter GlcP, PTS system; trehalose-specific IIA component (EC 2.7.1.69), PTS system; trehalsoe-specific IIB component, PTS system; trehalose-specific IIC component, Trehalose operon transcriptional repressor, trehalose-6-phosphate hydrolase. Another subsystem which was found to be present in NCTC 10301 but not MG1 is G3E family of P-loop GTPases and the role of genes in the operon are Urease accessory protein UreD, E, E, G, Urease Alpha subunit, beta subunit, and gamma subunit.

The MG1 subsystem was compared with the type strain of *Actinomyces oris*-CCUG 33920 and the genes, which were found in CCUG 33920 versus MG1, are mentioned in Appendix 3B. L-ascorbate utilization (and related gene clusters) subsystem of category carbohydrates was found to contain 3 genes, the role of which were 3-keto-L-gulonate 6-phosphate decarboxylase homolog, L-xylulose 5-phosphate 3-epimerase homolog, Putative 2-keto-3-deoxygluconate kinase. The genes for inositol catabolism were Glyceraldehyde-3-phosphate ketol-isomerase and inosose isomerise. There were three genes found which were involved in DNA Metabolism of subsystem Restriction-Modification System and these were Type I restriction-modification system, DNA-methyltransferase subunit M, restriction subunit R and specificity subunit S.

The subsystems of type strains of NCTC 10301 (AN, absent) versus CCUG 33920 (present, AO) were observed (Appendix 3C) and it was found that Lacto-*N*-Biose I and Galacto-*N*-Biose metabolic pathways were present in CCUG 33920 (AO) and four genes were found to play a role in Lacto-*N*-Biose utilization and uptake. These

genes were Lacto-*N*-Biose phosphorylase, predicted galacto-N-Biose-/lacto-N-Biose I ABC transporter, periplasmic substrate-binding protein; Predicted galacto-N-Biose-/lacto-N-Biose I ABC transporter, permease component 1; Predicted galacto-N-Biose-/lacto-N-Biose I ABC transporter, permease component 2. Another interesting Subsystem ABC transporter oligopeptide have three genes Oligopeptide transport ATP-binding protein OppF, Oligopeptide transport system permease protein OppB, Oligopeptide transport system permease protein OppC (Appendix 3C).

The genes which were unique for NCTC 10301 (*A. naeslundii*) but absent in CCUG 33920 (*A. oris*) are shown in Appendix 3D. Trehalose uptake and utilization subsystem was found to contain a set of 6 genes and the G3E family of P-loop GTPases which are involved in urease transport contained 7 Urease accessory genes.

Chapter 4 COMPARATIVE GENOMICS OF

A. oris & *A. naeslundii*

4.0 Comparative Genomics

The field of genome sequencing of bacteria has been progressing enormously since the first bacterial genome, *Haemophilus influenzae*, was sequenced by Fleischmann and colleagues in 1995 (Fleischmann *et al.*, 1995). More than 6342 bacterial genomes have been sequenced and made publically available and nearly 1589 projects are ongoing (http://www.genomesonline.org). The field of genomics is one of the most remarkable developments in science over the past decade.

Initially when the first two bacterial genomes were subject to sequencing, a question was raised that a comparison of the two genomes may be a sign of the minimal gene set required for bacterial existence (Mushegian & Koonin, 1996). The bacterial biology was studied in two strains which were non-infectious. These strains were very small and not closely related therefore a simple life style was estimated. However, it is obvious that to survive in different environmental niches, additional gene sets were required. The hypothesis is that when more and more genome sequence became available then it would be possible to say "one size fits all" in the entire bacterial domain (Pallen & Wren, 2007).

Today, with hundreds of genomes available on the internet, questions have been raised as to what we have learnt from bacterial genomics? On the very first hand, the continuous development of genomic data provides the basis for various applications of comparative genomics. The use of comparative genomics is useful as any single genome can provide insight into only that given organism while the comparisons of multiple genomes provide substantially more information on the physiology and evolution of bacteria. Better functional annotations can be given for the predicted coding sequences and new genes were identified (Fraser *et al.*, 2000). Comparative genomics is a powerful tool in revealing bacterial genomic diversity and sheds light on genomic variation in closely related species. The technologies for comparing bacteria at genomic, transcriptiomic and proteomic levels in recent years have been extensively reviewed (Binnewies *et al.*, 2006). Two methods are summarised here.

4.1 Multi Locus Sequence Typing (MLST)

MLST was conducted in order to determine the phylogenetic position of additional *A. oris* and *A. naeslundii* isolates from the internet. The sequence data from MLST and data from the internet for c505, MG1, OT170, OT171, OT175 and *A. johnsonii* was concatenated and a new Neighbor-Joining tree was calculated.

4.1.1 Material And Methods

4.1.1.1 Concatenation Of House Keeping Genes And Phylogeny

Seven housekeeping genes were chosen for MLST analysis in order to distinguish reference strains of *A. naesludii* genospecies 1 and 2 (*A. naeslundii* and *A. oris*), *Actinomyces* serotype WVA 963 (*A. johnsonii*) and a collection of 100 clinical oral and non-oral strains. Partial DNA sequences of these genes were analyzed regarding their phylogenetic distances to each other. These house keeping genes were *atp*A (ATP synthase F1, alpha subunit, ANA_0169), *rpo*B (DNA-directed RNA polymerase, beta subunit, ANA_1497), *pgi* (glucose-6-phosphate isomerase, ANA_0727), *Met*G (methionyl-tRNA synthase, ANA_1898), *glt*A (citrate synthase I, ANA_1674), *gyr*A (DNA gyrase, subunit A, ANA_2224)], *phe*S (Phenylalanyl-tRNA synthetase, alpha subunit) GQ354571-354683) were identified from the genome of *A. naeslundii* MG1. The following website http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi was used to extract the sequences of housekeeping genes. These genes were selected as they were present as single copies in the *A. oris*-MG1 genome, were widely spaced on the chromosome and were of sufficient size for primer design to yield amplicons of 450 bp. The 69 *A. oris* isolates and 31 *A. naeslundii* isolates were selected (Henssge, 2009). The sequences were kept in a separate file for the corresponding genes. The MEGA software version 4.0 was used to create the Neighbor-Joining tree and 7 house keeping genes of OT171, OT170, OT175, MG1, c505 and K20 isolates were included in the

MLST analysis. *A. johnsonii* could not be included in MLST analysis as the *pheS* gene was not available in the publically available database. The DNA sequences of individual genes were aligned using ClustalW Multiple Alignment with the FAST algorithm for guide tree (Thompson, 1994) and saved as a fasta file. Phylogenetic calculations of these alignments were conducted using MEGA 3.1 software (Kumar, 2004). Phylogenetic reconstruction was done using Bootstrap Neighbor-Joining method with 500 replicates and Kimura 2-parameter for nucleotides. Resulting phylogenetic trees were viewed and edited with the tree explorer in MEGA 3.1. Bootstrap values specify the reliability of a branching point by showing the percentage of times the same node appears in the 500 tree replicates. A bootstrap value of 95% is considered as giving a reliable branching point (Felsenstein, 1985).

4.1.2 Results

4.1.2.1 Population Diversity/Structure

The population structure of *A. oris* (n=69) and *A. naeslundii* (n=37) isolates with MG1, k20, c505, OT170, OT171, OT175 were analyzed individually. Separate alignment fasta files were prepared for each house keeping gene. The MLST analysis results revealed that the majority of *A. oris* (n=56) isolates cluster homogeneously. S24V, R21091 and S64C seem to cluster slightly differently from the main *A. oris* cluster but they are close to it. OT175 falls within the clusters of *A. oris* isolates. OT171, P6N and CCUG 33920 came out as separate branches between *A. oris* and *A. naeslundii*. OT170 falls in the middle of *A. oris* and *A. naeslundii* clusters (Figure 4.1). Some general finding was observed using DNA sequences of these seven house-keeping genes that OT175, MG1, k20, c505 (marked ▲ in Figure 4.1) resembled to *A. oris* groups in all genes which were already concatenated during the previous study (Henssge *et al.*, 2009). OT171 (■) and OT170 (●) formed distinct branches, while all (n=37) *A. naeslundii* isolates clearly formed a separate cluster and they all are distinct from *A. oris*

(Figure 4.1). The strains which were outgroup in Figure 3.1 and 3.2 are marked out with color symbols in Figure 4.1. These genes represent the most promising genes for genospecies discrimination as they revealed sufficient inter-species differences as well as high intra-species homogeneity. The MLST analysis of *A. oris* and *A. naeslundii* isolates showed a clonal structure in each population.

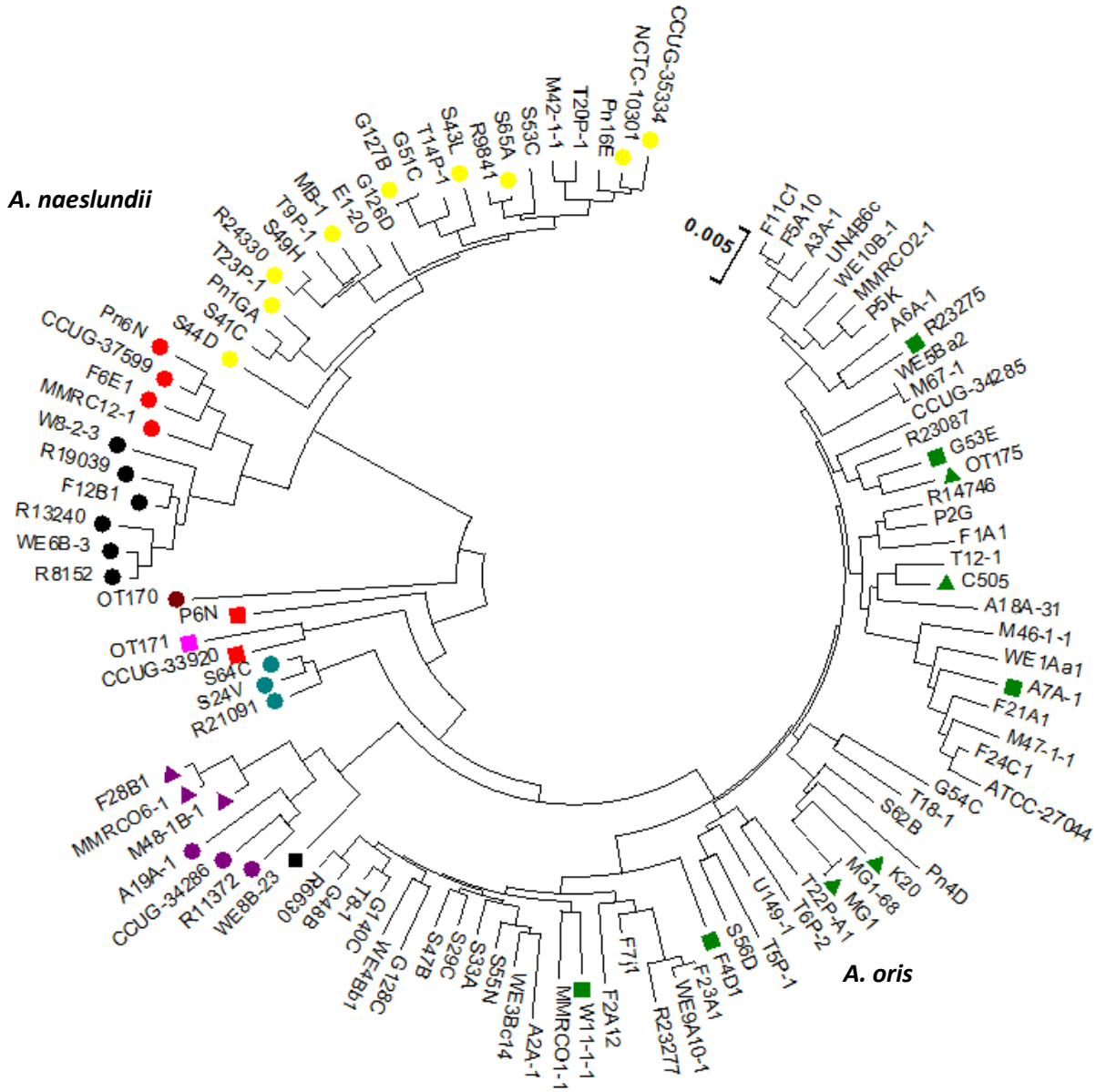


Figure 4.1: Concatenation of 7 house keeping genes.

A. oris **: ■ ; *A. oris*_1 :■ ; *A. oris*_2 :● ; *A. oris*_3:◀ ; *A. oris*_4: ■ ;**

***A. oris*_5: ● ; *A. naeslundii*: ● ; *A. naeslundii*_1: ● ; *A. naeslundii*_2: ●**

***[c505, MG1, k20, OT175 (*A. oris*): ▲ ; OT170: ● ; OT171: ■ (data obtained from HOMD for MLST analysis)]**

*****A. johnsonii* was not included in MLST analysis as it lacks pheS gene.**

4.2 Digital DNA-DNA Hybridization

4.2.1 Introduction

Generally it was accepted by the scientific community that taxonomic information lies in complete nucleotide sequence/whole genome of a bacterium (Goris *et al.*, 2007; Stackebrandt *et al.*, 2002; Wayne *et al.*, 1987). The realistic approach for defining the species of bacteria lies in finding DNA-DNA hybridization (DDH) or homology among genomes. Since the 1960's the DNA-DNA homology values were obtained using wet-laboratory techniques which were the most authentic genome-wide comparisons among bacterial organisms at that time. 70% similarity was set for delineating bacterial species by Wayne *et al*. (1987). Taxonomists used several wet-laboratory techniques to obtain an approximate value of similarity among genomes of two strains but the techniques adopted were so tedious, error-prone and unfortunately could not help much to build up a comparative database. The large amount and high quality DNA was also required to implement the DDH technique in the laboratory as compared to PCR-based studies. The availability of diverse methods were the basis of variable results (Grimont, 1980; Huss *et al.*, 1983). These were the drawbacks and reasons why the research community was always looking for alternative and cheap methods to replace tedious DDH experiments (Cho & Tiedje, 2001; Coenye *et al.*, 2005; Gevers *et al.*, 2005; Stackebrandt, 2003). Recently, the advancement of gene-sequencing techniques by using next generation whole genome sequencing approaches has facilitated the comparative analysis and several bioinformatic methods were developed to replace the wet-laboratory DDH by in-silico genome-to-genome comparisons. Digital DDH have revolutionised comparative genomics since the technology was first used for comparing several strains (Goris *et al.*, 2007). Since then, DDH technology has been widely used for comparative analysis of the genome content and genetic variation of different bacterial species. Digital DDH has become one of the commonest methods to quickly reveal the gene contents of many closely related strains of a species. This approach is referred to as Comparative genome hybridization analysis (CGH). This technique can therefore provide a rapid assessment

of the differences between two genomes. DDH calculation was also considered as a major tool for the delineation of microbial species (Auch *et al.,* 2010b). Furthermore, the digital DDH values were determined in the current study among related strains of *Actinomyces* for which the whole genome has been sequenced. Digital DDH has been used in studies by Konstantinidis *et al*., (2005), Deloger *et al* (2009), Meier-Kolthoff *et al* (2013) and by Auch *et al* (2010a).

4.2.2 Material And Methods

DDH studies of this kind determine the homology of all the common genes of a sequenced and annotated reference strain. DDH estimates were calculated using GGDC (Genome-to-Genome Distance calculator) version 2.0 (http://ggdc.dsmz.de), an online calculator using the strategy of GBDP (Genome Blast Distance Phylogeny). Data were obtained from whole genome sequences of *Actinomyces* strains using next generation sequencing technologies of Illumina Genome Analyzer and Roche 454. The goal of the present study was to examine more accurately the relationship between DDH values obtained from whole genomic sequences of *A. oris* and *A. naeslundii* isolates.

Sets of strains were selected from two groups of *Actinomyces* (*A. oris* and *A. naeslundii*) for the comparison. Seventeen named strains of *A. oris* and nineteen strains of *A. naeslundii* were investigated by nucleic acid hybridization. In addition to the isolate collection of *A. naeslundii/oris,* seven other oral *Actinomyces* spp. were included in the current study of sequence-based DDH calculations. These strains were MG1 (SEQF1063), k20 (SEQF2431), OT170 (SEQF1849), OT171 (SEQF1693), OT175 (SEQF1848), *A. johnsonii* (SEQF1668) and *A. odontolyticus* (SEQF1388) as their genomic sequences were publically available (www.homd.org ((Chen *et al.*, 2010)). The final data set comprised 43 genomes including genomes from the internet database listed above. Strains in this study were chosen on the basis of the availability of their

complete genome sequences (either fully closed or high draft sequences) and as their 16S sequences exhibited high level of similarity to these of *A. oris* and *A. naeslundii* (Henssge *et al.*, 2009). The strains listed in Table 2.1 were provided from the microbiology lab of King's College London. The alignment method used for finding inter-genomic distances was NCBI-BLAST. The FASTA files used were those created during denovo analysis (Section 3.2.1). The one FASTA file was selected for the query genome and up to 10 FASTA files were selected in the reference genomes list. Each FASTA query and reference genomes were given a name. The email address was provided to get the results. The online submission form compares query genome with reference or known genome and used three distance formulas to calculate the distance among genomes. The results obtained with Formula 2 were considered in this study as suggested by Auch (Auch *et al.*, 2010b). Formula 2 (Identities/HSP length (High-scoring segment pairs) is a stronger option to use against the use of incomplete draft genome and therefore the results obtained were independent of genome length. The calculated DDH estimate was regression based. The predicted value closer to 70% is of interest particularly and regression (with a special type of Generalized linear model (GLM)) is used for reporting the DDH is >=70%. The jobs submitted took several minutes to process due to workload on the server or job size. The DDH estimates produced using the computational method in the current study is both accurate and precise. The flow chart given in Figure 4.2 describes the steps involved for digital DNA-DNA Hybridization calculation.

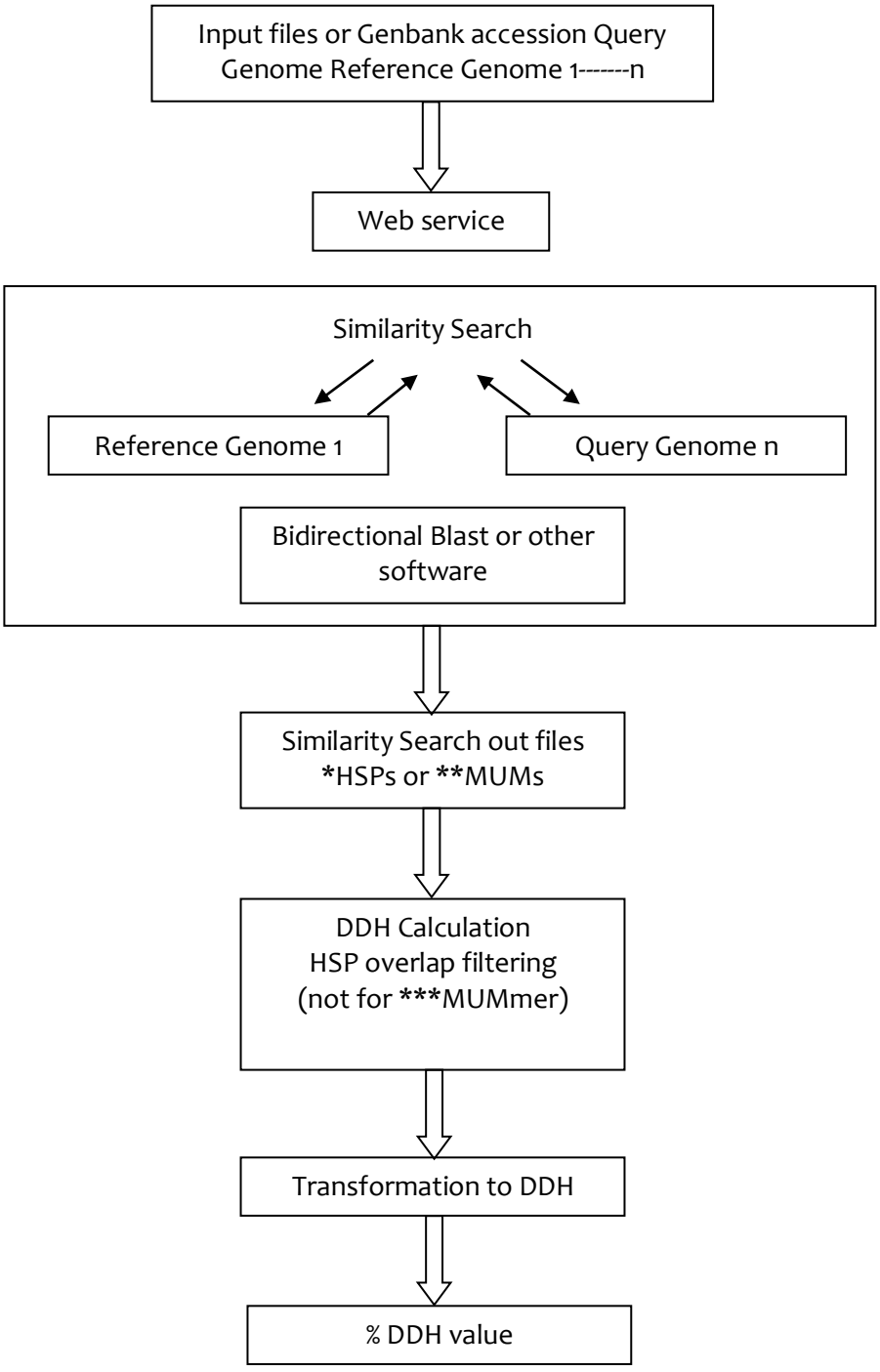


Figure 4.2: A flow chart to calculate DDH values in silico.

***FASTA Files or GenBank files were uploaded in the server and final values were received through email (Adapted from Auch *et al* 2010b). * HSP: High-scoring segment pairs ** MUMs (Maximally unique Matches) *** MUMmer : Software package to align large scale DNA sequences.**

4.2.3 Results

The DDH values were measured between the six *A. oris* and 3 *A. naeslundii* bacterial groups and are presented in Tables 4.1 and 4.2. The sequences of all other isolates of MG1, k20, c505, OT170, OT171, OT175, *A. johnsonii* and *A. odontolyticus* were available on the internet. The DDH value of 100% was determined when the same strain was tested against itself and when compared to test species of *A. johnsonii* and *A. odontolyticus*, the DDH was 35% and 23% respectively. The DDH values of the *A. oris* cluster (A7A-1, F4D1, G53E, R23275, and W11-1-1) along with internet strains OT175, c505, MG1 and k20 were compared among themselves and with each other as shown in Table 4.1 (highlighted yellow in Table 4.1). The homology values shown were 100% for the *A. oris* strains when compared among themselves e.g A7A versus A7A, F4D1 versus F4D1, G53E versus G53E, R23275 versus R23275, while 53.3 - 75.5% homology values were observed for the *A. oris* strains when tested with each other e.g A7A versus F4D1 had 60.1 % homology values and all other values were shown in Table 4.1. The DNA-DNA relatedness values of *A. oris* strains were also compared with MG1, which showed on average 56% DNA homology (Table 4.1). OT175 and G53E, k20 and MG1 and W11-1-1 and F4D1 form the same species but all other *A. oris* are different and have DDH < 70% therefore the results obtained showed that most of the strains in the *A. oris* group are distinct and separate from each other.

Similarly when all members of the *A. oris* (AO) cluster were compared with members of clusters AO_1 to AO_5, the average homology values observed for AO versus AO_1 were 48.8% ± 0.29, AO_2 showed 47.5% ± 0.38, AO_3 gave 48.3% ± 0.30, AO4_A and AO4_B showed 43.0% ± 0.24 and AO_5 had 47.4% ± 0.28 (Table 4.1). The DDH values of two strains of cluster AO_4A and AO_4B, named P6N and CCUG 33920 respectively, was very low as compared to other strains of AO clusters, which showed higher DDH values. In summary, the DDH values of AO_1 to AO_5 clusters were in the range of 42.7-49.2% when compared to the other AO strains therefore all these clusters are different to the main AO cluster (Table 4.1).

The DNA relatedness was observed for AO_1 to 3 when compared to each other, the estimate was in the range of 71.90 to 100 % (DDH values shown in red in Table 4.1). The probability of being the same species is very high. When AO_1 was compared with AO_2 and AO_3, they showed average homology values of 72.7 and 74.8% respectively, while with clusters AO_4A and AO_4B showed only 42.8% and 43.2%, however with AO_5 it had 51.2% and finally with MG1 showed 48% (Table 4.1). AO_2 versus AO_3 had average homology (72.8% ± 0.21), AO_4A, 4B showed average homology (43% ± 0.16), and AO_5 had (51.4% ± 0.25) and finally with MG1 had average homology (47% ± 0.13). AO_3 versus AO_4A, AO_4B had (43% ± 0.07), with AO_5 had (51.0% ± 0.03) while with MG1 had (47% ± 0.11). AO_4 gave 45% ± 0.05 homology with AO5 (Table 4.1). Therefore, the clusters of AO_1, 2 and 3 are identical while clusters of AO_4 and AO_5 are distinct from each other.

The DDH values of >=70% were observed among strains of cluster AO_5 (S64C, R21091 and S24V) (presented green in Table 4.1) therefore the strains of cluster AO_5 are the same species. Similarly a couple of DDH values of >=70% were observed, these are CCUG 33920 and OT171 (84.9%), MG1 and k20 (79.4%), W11-1-1 and F4D1 (75.5%) and G53E and OT175 therefore they are all same species (Table 4.1, highlighted red).

The strains which were sequenced belonging to the true AN cluster showed average DNA homology values in the range of 81.6-100% (showed red in Table 4.2) and there no significant difference was observed when compared with each other (Table 4.2). The average homology value was in the range of (43.14 % ± 0.44) when tested with MG1. Similarly, the strains of AN_1 gave 69-71% average DNA homology values when compared to the true AN cluster (showed green in Table 4.2). The selected strains belonging to AN cluster 1 and 2 were also compared among themselves and with each other and found to have 85-100% homologies when AN_1 was compared with AN_1

(values shown in green in Table 4.2) and 73-75% homology values when compared with AN_2 (shown blue in Table 4.2). The DDH values of 87-100% were observed when AN_2 tested with AN_2 (shown pink in Table 4.2). In conclusion, all the DDH values were in the range of 70 % – 100 % therefore all the strains of *A. naeslundii* belong to a single cluster.

The DDH values were also observed when all strains of cluster AO to AO_5 were tested for homology with AN to AN_2. The DDH values of all of them fall in the range of 35-36% therefore *A. oris* clusters were clearly a separate group from *A. naeslundii* clusters (Table 4.3). The same finding was observed when tested with *A. johnsonii* and *A. odontolyticus*. They both are distinct to all other *A. oris* and *A. naeslundii* clusters. The DDH findings are in agreement with the MLST analysis results which also showed similar groupings/clusters based on house-keeping gene analysis.

Table 4.1: Digital DDH values of *A. oris* versus *A. oris* (Table is continuous on next page)

Name of Organisms	G53E	OT175	W11-1-1	F4D1	R23275	A7A-1	C505	MG1	K20	WE8B-23	R11372	A19A-1
G53E (AO)	100	74.0	60.4	60.8	67.7	66.6	67.7	56.5	56.9	48.3	47.2	47.3
OT175 (AO)	74.0	100	60.7	60.6	68.6	67.2	68.8	56.4	56.6	48.7	47.4	47.4
W11-1-1 (AO)	60.4	60.7	100	75.5	60.5	60.0	63.5	53.3	54.0	48.7	47.5	47.6
F4D1 (AO)	60.8	60.6	75.5	100	60.3	60.1	64.2	53.9	54.7	48.7	47.3	47.4
R23275 (AO)	67.7	68.6	60.5	60.3	100	69.6	68.1	56.7	57.0	49.2	48.2	48.2
A7A-1 (AO)	66.6	67.2	60.0	60.1	69.6	100	68.5	56.4	57.0	49.0	48.0	47.9
c505 (AO)	67.7	68.8	63.5	64.2	68.1	68.5	100	56.2	56.5	49.1	48.0	47.7
MG1	56.5	56.4	53.3	53.9	56.7	56.4	56.2	100	79.4	48.6	47.0	47.1
k20 (MT 500 seq)	56.9	56.6	54.0	54.7	57.0	57.0	56.5	79.4	100	49.1	47.7	47.8
WE8B-23 (AO_1)	48.3	48.7	48.7	48.7	49.2	49.0	49.1	48.6	49.1	100	71.9	73.9
R11372 (AO_2)	47.2	47.3	47.5	47.3	48.2	48.0	48.0	47.0	47.7	71.9	100	79.1
A19A-1 (AO_2)	47.3	47.4	47.6	47.4	48.2	47.9	47.7	47.1	47.8	73.9	79.1	100
CCUG 34286(AO_2)	47.2	47.9	47.4	47.2	48.3	48.0	47.5	46.9	47.6	72.3	87.4	79.0
F28B1 (AO_3)	48.1	48.3	48.5	48.3	48.8	48.6	48.3	48.0	48.6	74.3	72.6	73.0
M48-1B-1 (AO_3)	48.0	48.3	48.3	48.1	48.9	48.8	48.3	47.9	48.3	75.5	72.5	73.6
MMRCO6-1 (AO_3)	48.1	48.2	48.3	47.8	48.8	48.7	48.4	47.8	48.2	74.5	72.3	73.2
P6N (AO_4A)	42.7	42.9	42.8	42.7	42.9	42.8	42.8	42.5	42.9	42.8	42.5	42.6
CCUG 33920(AO_4B)	43.2	43.4	43.0	42.9	43.3	43.0	43.1	42.6	43.3	43.2	43.1	43.1
OT171(MT 500 seq)	43.4	43.8	43.4	43.3	43.7	43.4	43.3	43.0	43.1	43.4	43.0	43.2
S24V (AO_5)	47.6	47.4	47.7	47.4	47.4	47.6	47.7	46.9	47.2	51.3	51.5	51.7
R21091 (AO_5)	47.0	47.2	47.4	47.3	47.4	47.3	47.3	46.7	47.3	51.1	51.1	51.4
S64C (AO_5)	47.5	47.1	47.4	47.2	47.1	47.5	47.5	46.6	46.9	51.2	51.4	51.6
OT170	34.9	35.1	34.9	34.9	35.0	35.1	35.1	34.6	35.2	35.3	35.1	35.2
A. johnsonii	35.4	35.5	35.3	35.3	35.6	35.5	35.5	35.0	35.5	35.8	35.5	35.5
A. odontolyticus	23.8	23.8	23.7	23.8	23.7	23.7	23.8	23.7	24.0	23.6	23.5	23.5

CCUG 34286	F28B1	M48-1B-1	MMRCO6-1	P6N	CCUG 33920	OT171	S24V	R21091	S64C	OT170	A. j	A. od
47.2	48.1	48.0	48.1	42.7	43.2	43.4	47.6	47.0	47.5	34.9	35.4	23.8
47.9	48.3	48.3	48.2	42.9	43.4	43.8	47.4	47.2	47.1	35.1	35.5	23.8
47.4	48.5	48.3	48.3	42.8	43.0	43.4	47.7	47.4	47.4	34.9	35.3	23.7
47.2	48.3	48.1	47.8	42.7	42.9	43.3	47.4	47.3	47.2	34.9	35.3	23.8
48.3	48.8	48.9	48.8	42.9	43.3	43.7	47.4	47.4	47.1	35.0	35.6	23.7
48.0	48.6	48.8	48.7	42.8	43.0	43.4	47.6	47.3	47.5	35.1	35.5	23.7
47.5	48.3	47.9	47.8	42.8	43.1	43.0	46.9	47.3	46.6	34.6	35.0	23.7
46.9	48.0	48.3	48.4	42.5	42.6	43.3	47.7	46.7	47.5	35.1	35.5	23.8
47.6	48.6	48.3	48.2	42.9	43.1	43.3	47.2	47.3	46.9	35.2	35.5	24.0
72.3	74.3	75.5	74.7	42.8	43.2	43.4	51.3	51.1	51.2	35.3	35.8	23.6
87.4	72.6	72.5	72.3	42.5	43.1	43.0	51.5	51.1	51.4	35.1	35.5	23.5
79.0	73.0	73.6	73.2	42.6	43.1	43.2	51.7	51.4	51.6	35.2	35.5	23.5
100	73.1	72.9	72.6	42.5	42.9	43.1	51.5	51.1	51.4	35.1	35.5	23.4
73.1	100	91.0	88.1	43.0	43.1	43.3	51.1	51.0	51.1	35.4	35.7	23.7
72.9	91.0	100	86.8	43.1	43.0	43.4	51.1	50.9	50.9	35.4	35.8	23.7
72.6	88.1	86.8	100	42.8	43.0	43.3	51.2	50.9	51.1	35.3	35.7	23.4
42.5	43.0	43.1	42.8	100	47.5	47.9	44.8	44.5	44.6	35.6	36.0	23.6
42.9	43.1	43.4	43.3	47.5	100	84.9	44.9	44.5	44.6	36.0	35.8	24.2
43.1	43.3	43	43.0	47.9	84.9	100	44.7	44.6	44.5	35.8	35.9	23.5
51.5	51.1	51.1	51.2	44.8	44.7	44.9	100	80.1	87.2	35.8	36.6	23.8
51.1	51.0	50.9	50.9	44.5	44.5	44.6	80.1	100	80.0	35.4	36.1	23.3
51.4	51.1	50.9	51.1	44.6	44.5	44.6	87.2	80.0	100	35.5	36.4	23.5
35.1	35.4	35.4	35.3	35.6	35.8	36.0	35.8	35.4	35.5	100	37.7	23.4
35.5	35.8	35.8	35.7	36.0	35.9	35.8	36.6	36.1	36.4	37.7	100	24.2
23.4	23.7	23.7	23.4	23.6	23.5	24.2	23.8	23.3	23.5	23.4	24.2	100

***** Yellow highlights indicates DDH values for "*A. oris*" group
****** Red highlights indicates the DDH values are more than 70% and known as same species
*******A. j: *A. johnsonii*; A. od: *A. odontolyticus*

Table 4.2: Digital DDH values of *A. naeslundii* versus *A. naeslundii, A. odontolyticus* and *A. johnsonii*

Name of Organism	CCUG-35334	G127B	MB-1	NCTC_10301	R24330	S43L	S44D	S65A	T23P-1
CCUG 35334 (AN)	100	84.7	84.6	86.5	87.2	84.6	82.4	86.6	82.3
G127B (AN)	84.7	100	85.9	87.4	86.9	91.4	83.0	87.2	82.7
MB-1 (AN)	84.6	85.9	100	86.0	87.1	85.2	82.2	86.2	81.7
NCTC_10301 (AN)	86.5	87.4	86.0	100	90.3	88.6	83.5	89.2	82.4
R24330 (AN)	87.2	86.9	87.1	90.3	100	86.7	83.5	89.8	82.0
S43L (AN)	84.6	91.4	85.2	88.6	86.7	100	83.7	86.9	82.4
S44D (AN)	82.4	83.0	82.2	83.5	83.5	83.7	100	82.9	81.6
S65A (AN)	86.6	87.2	86.2	89.2	89.8	86.9	82.9	100	81.6
T23P-1 (AN)	82.3	82.7	81.7	82.4	82.0	82.4	81.6	81.6	100
CCUG 37599 (AN_1)	69.5	69.8	70.0	70.0	70.5	69.9	70.6	69.8	69.0
F6E1 (AN_1)	70.8	70.2	70.9	71.0	71.7	70.5	71.2	70.8	70.5
Pn6N (AN_1)	69.1	69.3	69.6	69.7	70.1	69.6	70.4	69.2	69.1
MMRC12-1 (AN_1)	69.1	69.3	69.6	69.7	70.1	69.6	70.4	69.2	69.1
F12B1 (AN_2)	70.8	70.7	70.7	71.0	71.2	71.2	72.0	70.9	70.6
R8152 (AN_2)	71.2	71.3	71.3	71.8	71.8	71.8	72.4	71.1	71.0
R13240 (AN_2)	71.4	71.8	71.6	72.2	72.4	72.4	72.3	71.7	71.1
R19039 (AN_2)	71.3	71.5	71.3	71.9	72.2	71.9	72.7	71.9	71.3
W8-2-3 (AN_2)	70.7	70.7	70.9	71.4	71.4	71.1	72.1	70.8	70.4
WE6B-3 (AN_2)	71.0	71.2	71.0	71.3	71.5	71.7	72.2	71.2	70.9
OT170	37.3	37.4	37.0	37.4	37.4	37.3	37.4	37.3	37.1
OT171 (MT 500 seq)	36.7	36.7	37.2	36.7	36.7	36.5	36.7	36.8	36.7
OT175	35.9	35.8	36.1	35.9	35.9	35.8	36.0	35.9	35.8
C505	36.0	35.8	36.1	36.0	35.8	35.8	36.1	35.9	35.7
K20 (MT 500 seq)	35.9	35.9	35.9	35.3	35.7	35.9	35.8	35.8	35.7
A. johnsonii	38.7	38.6	38.7	38.6	38.8	38.6	38.8	38.6	38.5
A. odontolyticus	23.6	23.6	23.6	23.6	23.6	23.7	23.8	23.5	23.7

CCUG 37599	F6E1	Pn6N	MMRC12-1	F12B1	R8152	R13240	R19039	W8-2-3	WE6B-3
69.5	70.8	69.1	69.1	70.8	71.2	71.4	71.3	70.7	71.0
69.8	70.2	69.3	69.3	70.7	71.3	71.8	71.5	70.7	71.2
70.0	70.9	69.6	69.6	70.7	71.3	71.6	71.3	70.9	71.0
70.0	71.0	69.7	70.7	71.0	71.8	72.2	71.9	71.4	71.3
70.5	71.7	70.1	70.1	71.2	71.8	72.4	72.2	71.4	71.5
69.9	70.5	69.6	69.6	71.2	71.8	72.4	72.2	71.4	71.7
70.6	71.2	70.4	70.4	72.0	72.4	72.3	72.7	72.1	72.2
69.8	70.8	69.2	69.2	70.9	71.7	71.7	71.9	70.8	71.2
69.0	70.5	69.1	69.1	70.6	71.0	71.1	71.3	70.4	70.9
100	86.3	88.1	85.4	73.2	73.0	72.5	72.5	72.4	73.5
86.3	100	86.1	93.6	73.6	73.5	73.1	73.0	72.9	74.2
88.1	86.1	100	85.4	72.6	72.1	72.0	72.0	71.8	72.9
85.4	93.6	85.4	100	75.8	75.1	74.5	75.3	74.4	75.5
73.2	73.6	72.6	75.8	100	88.4	90	98.9	89.6	89.7
73.0	73.5	72.1	75.1	88.4	100	87.3	88.0	88.7	88.8
72.5	73.1	72.0	74.5	90.0	87.3	100	89.2	90.3	90.4
72.5	73.0	72.0	75.3	98.9	88	89.2	100	88.8	89.1
72.4	72.9	71.8	74.4	89.6	88.7	90.3	88.8	100	92.0
73.5	74.2	72.9	75.5	89.7	98.4	90.4	89.1	92.0	100
37.0	37.0	37.1	36.8	37.1	37.1	37.0	37.1	37.1	37.1
35.9	36.0	35.9	36.0	36.4	36.4	36.4	36.3	36.3	36.4
35.1	35.2	35.2	35.1	35.6	35.6	35.7	35.6	35.5	35.6
35.4	35.4	35.4	35.3	35.7	35.7	35.7	35.7	35.5	35.7
35.4	35.3	35.3	35.3	35.3	35.4	35.3	35.3	35.2	35.3
38.7	38.7	38.5	38.6	38.6	38.8	38.6	38.6	38.6	38.6
23.3	23.4	23.4	23.2	23.5	23.5	23.5	23.4	23.2	23.5

* DDH estimates are in %

** OT170, OT171, OT175, c505 and k20 are *A. oris* isolates

Table 4.3: Summary of Final DDH Calculation

	A. oris	AO1	AO2	A03	AO4A	AO4B	AO5	AN	AN1	AN2
A. oris	66.29 (53.3-100)	48.8 (48.7-49.2)	47.5 (47.0-48.3)	48.3 (47.8-48.9)	43.0 (42.5-42.8)	43.0 (42.6-43.4)	47.4 (47.7-46.9)	36 (35.7-36.0)	35 (35.1-35.4)	35.5 (35.4-35.9)
AO1		100*	72.7 (71.9-73.9)	74.8 (74.3-75.5)	42.8	43.2	51.2 (51.1-51.23	36 (35.8-36.3)	35.5 (35.5-35.6)	35.9 (35.8-36.0)
AO2			87.8 (79-100)	72.8 (72.6-73.6)	42.5 (42.5-42.6)	43.0 (42.9-43.1)	51.4 (51.1-51.7)	35.7 (35.9-36.0)	36 (35.0-35.9)	36 (35.7-35.9)
A03				92.4 (86.8-100)	42.9 (4.8-43.0)	43 (43.1-43.0)	51.0 (50.9-51.2)	36.7 (36.0-38.1)	36 (36.0-36.2)	36 (35.0-36.0)
AO4A					100*	47.5	44.6 (44.5-44.8)	38.1 (38-38.3)	37.2 (37.2-37.2)	37.2 (37.2-37.4)
AO4B						100*	44.7 (44.6-44.9)	36.6 (36.5-36.6)	35.9 (35.9-36)	36.2 (36.1-36.2)
AO5							88.2 (80-100)	36.7 (36.5-37)	36 (36.0-36.5)	36 (36.1-36.8)
AN								86.8 (81.6-100)	70 (69.1-71.5)	71 (70.0-72.4)
AN1									90.6 (85.4-100)	75 (72.1-75.1)
AN2										91.6 (87.3-100)

***** Final summary of DDH values showing average and also the minimum and maximum values in brackets.

** Only one strain available.

***AO_4A and AO_4B represent CCUG 33920 and P6N isolates respectively

4.3 Comparative Analysis Of Whole Genomes Using Mauve

4.3.1 Material and Methods

Mauve version 2.3.1 software was downloaded from website (http://gel.ahabs.wisc.edu/mauve/) for multiple genome alignment, to determine the sequence re-arrangement structures within genomes and to provide a comprehensive picture of genetic differences among closely related genomes of *Actinomyces*. The whole genomes (Table 2.1) were compared and aligned using Mauve (Darling *et al.*, 2008). The analysis revealed that 43 genomes were closely related in genetic features to each other. An overview of genome alignments is shown in Figure 4.3. The alignments in Figure 4.3 showed the locally collinear blocks (LCBs) which represent the conserved or homologous regions in all of the genomes investigated. LCB is defined as a "collinear set of exactly matched subsequence (multiple maximal unique matches namely 'multi-MUMS') which is shared by all the input sequences" (Chaves *et al.*, 2010; Darling *et al.*, 2008). The weight (the sum of the lengths of the included multi-MUMs) of a LCB serves as a measure of confidence that it is a true homologous region rather than a random match (Darling *et al.*, 2008). There are also white areas in the LCB region representing less similar sequences and shows much more diversities. The draft genome used in the analysis has contigs which were sorted with denovo assembly and aligned using MG1 as a reference genome. The initial set of anchors of minimum length of 2.9 Mbases consumed ~ 144h on a 1.6 GHz desktop workstation. Computing the anchors consumes roughly 4GB RAM. Figure 4.4 highlights only six *A. oris* and 6 *A. naeslundii*. The unique genomic regions resemblance is presented with a green color and absence with a purple color.

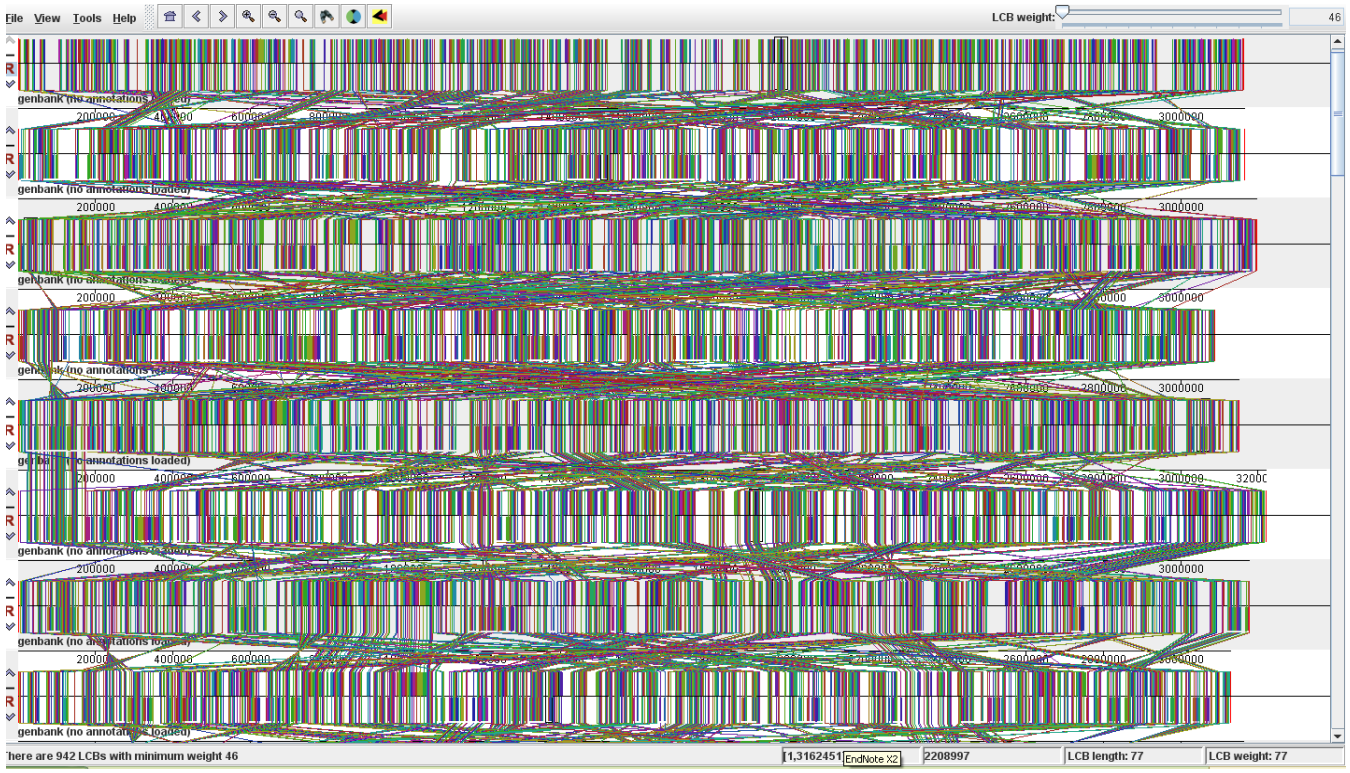
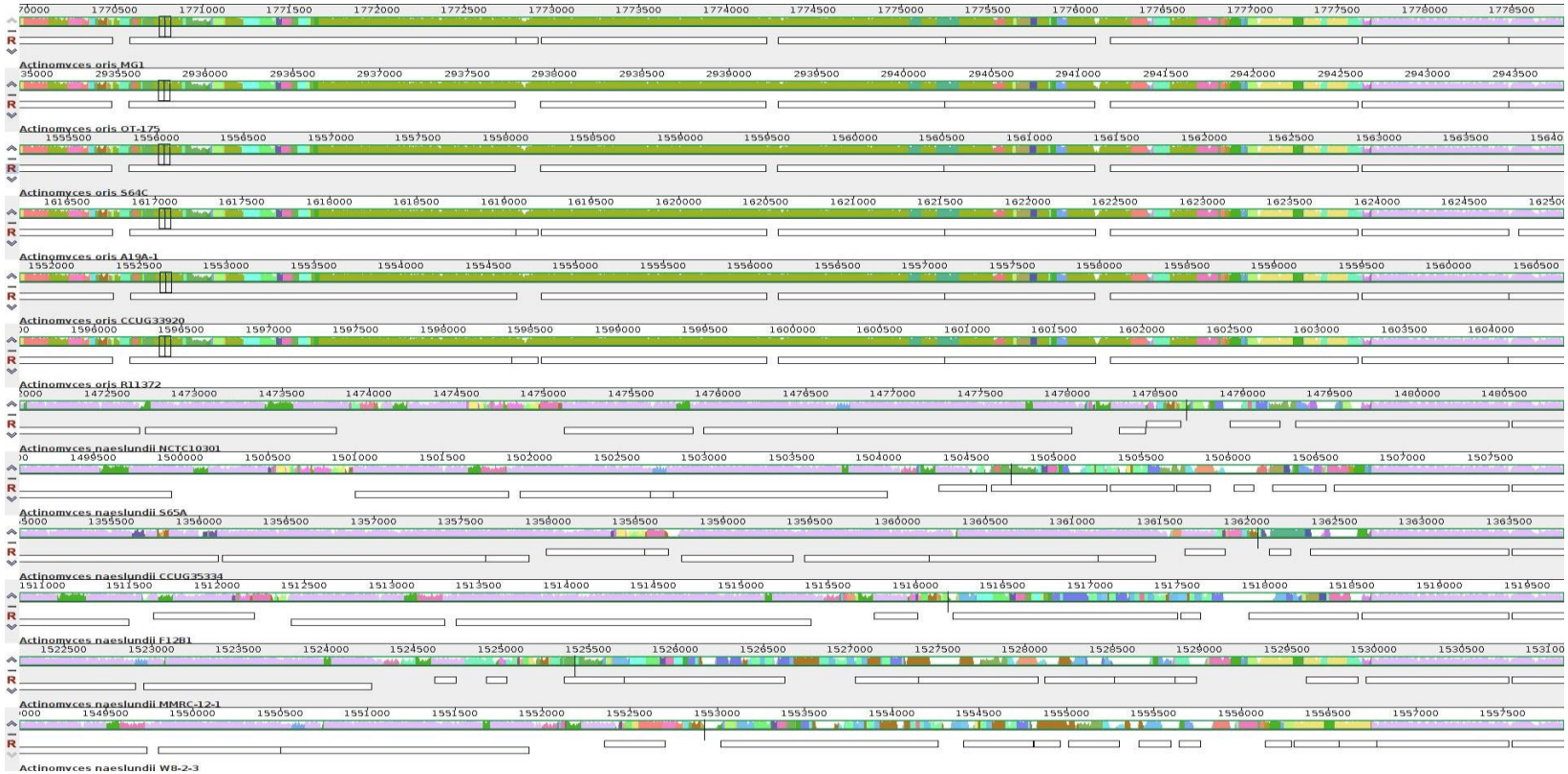


Figure 4.3: Mauve visualization of LCBs rearrangements in representative *Actinomyces* genomes.

***Each horizontal line represent single genome and vertical lines represent the homologus gene regions among genomes.**



A. oris

A. naeslundii

Figure 4.4: Locally Collinear blocks identified among the 43 *Actinomyces* genomes.

***Screenshot of lacto-*N*-biose operon region in 6 *A. oris* strains (MG1, OT175, S64C, A19A-1, CCUG 33920 and c505) and 6 *A. naeslundii* strains (R11372, NCTC-10301, S65A, CCUG35334, F12B1 and MMRC12-1) ; Unique genomic region gain shown in green and absence in purple colour**

4.3.2 Results Of Mauve Analysis

Genomes of *Actinomyces* were the most divergent and based on NJ-tree information (Figure 4.5), the isolates were divided into two groups. One group composed of *A. oris* has six subgroups/subspecies and the other consisted of *A. naeslundii* with three subgroups/subspecies. The *A. oris* cluster consisted of genomes of MG1, k20, F4D1, W11-1-1, G53E, OT175, c505, A7A-1, R23275; WE8B-23 is in cluster *A. oris*_1; A19A-1, CCUG 34286, R11372 included in the *A. oris*_2 cluster; MMRCO6, F28B1, M48-1B-1 is in *A. oris*_3 cluster; P6N, CCUG 33920 and OT171 grouped in *A. oris* 4 cluster; and finally R21091, S24V and S64C clustered as *A. oris*_5. The other main group consisted of *A. naeslundii* which have isolates clustered in the true *A. naeslundii* cluster named S44D, T23P-1, CCUG 35334, NCTC_10301, G127B, S43L, MB-1, R24330 and S65A. The cluster_1 of *A. naeslundii* consisted of isolates named F12B-1, R19039, R8152, R13240, W8-2-3, and WE6B-3 while cluster_2 consisted of isolates F6E1, MMRC-12-1, CCUG 37599 and Pn6N. OT170 and *A. johnsonii* were not associated with either group. Figure 4.5 analyses revealed consistent results as those obtained when compared to MLST data and digital DDH analysis.

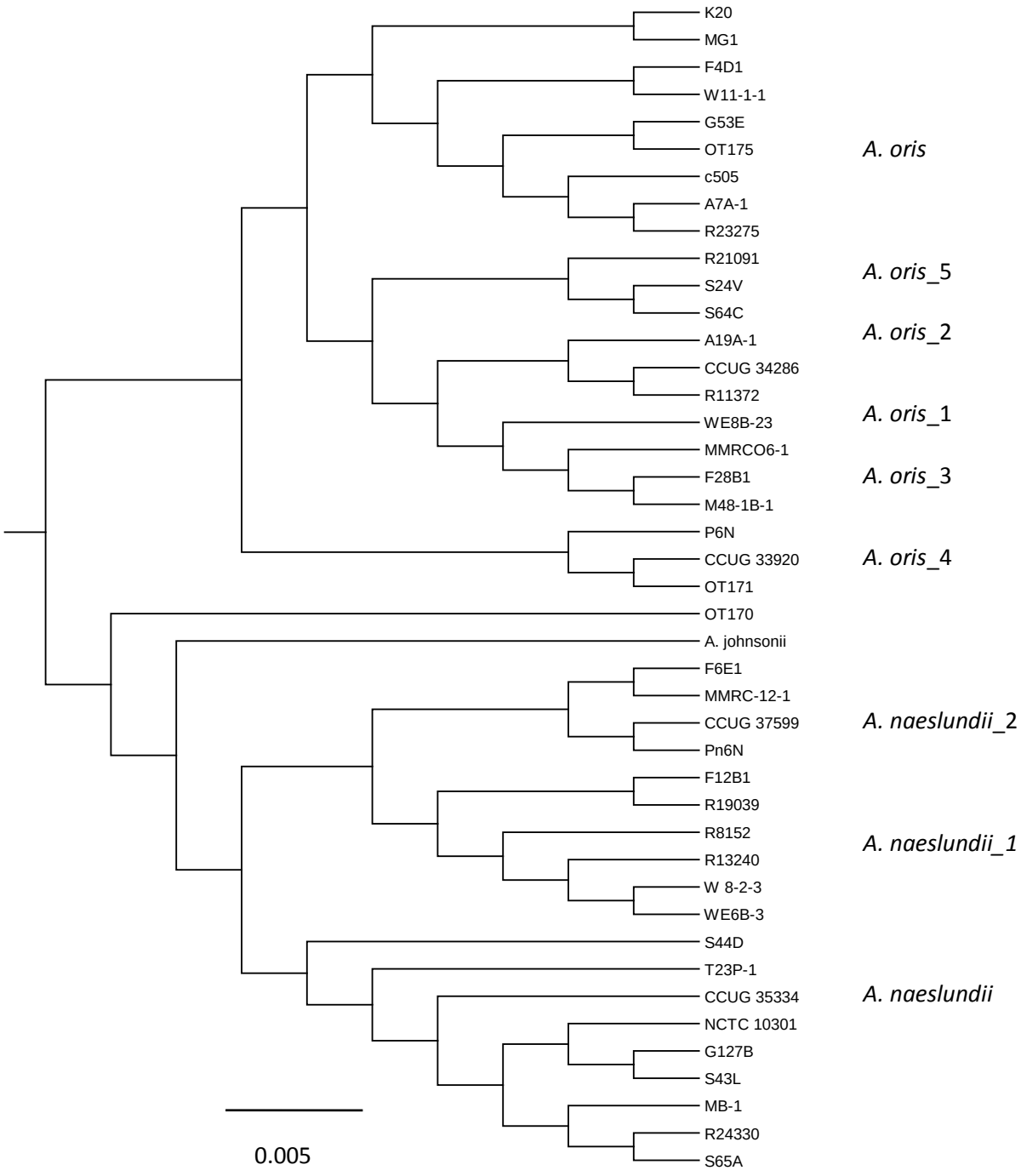


Figure 4.5: A Phylogenetic tree showing relationship after Mauve run of 43 *Actinomyces* genomes.

4.4 Comparative Analysis Using Progressive Mauve

4.4.1 Whole Genome Analysis

4.4.1.1 Material and Methods

Progressive mauve (PM) alignment tool (Darling *et al.*, 2010) was preferred to align the genomic sequences as it performs better than Mauve. The only minor disadvantage of PM is that it is a lot slower programme but also much more accurate than the original version of mauve. This programme is independent of genes, and works by using only the genome sequences to find which fragments are shared amongst which subset of genomes. The PM was used to align 43 *Actinomyces* genomes but the run repeatedly crashed even after running for more than 5 days. PM was applied to 17 *A. oris* isolates (Table 2.1) along with c505, MG1, k20, OT170, OT171, OT175 and 19 *A. naeslundii* strains separately as it was not possible to run 43 genomes using PM (Figure 4.6 & 4.7). PM worked for smaller data sets therefore it was concluded that PM worked perfectly fine with fewer number of genomes (less than 25). The **(**.backbone**)** file is an output file from PM which was used to run StripsubsetLCBs (http://gel.ahabs.wisc.edu/mauve/snapshots/). StripsubSetLCBs is software used to align the sequences created during the PM programme and converts them into exactly the format which was needed to run the ClonalFrame analysis programme (version 1.1). This identifies common sequences between strains. Ten independent runs of ClonalFrame were performed, each consisting of 40,000 iterations. Five hundred bootstrap replicates were generated. Furthermore a maximum likelihood tree was constructed for each individual gene using the MEGA programme (Figure 4.6 & 4.7).

4.4.1.2 Results

Six independent groups were observed known as *A. oris*, *A. oris*_1, *A. oris*_2, *A. oris*_3, *A. oris*_4, *A. oris*_5 when ClonalFrame analysis of 22 *A. oris* isolates was performed. Similarly when ClonalFrame analysis was performed on 19 *A. naeslundii* isolates, three independent groups were observed consisting of *A. naeslundii*, *A. naeslundii_1* and *A. naeslundii_2*. The same results were obtained as those using MLST, Digital-DDH and Mauve analysis.

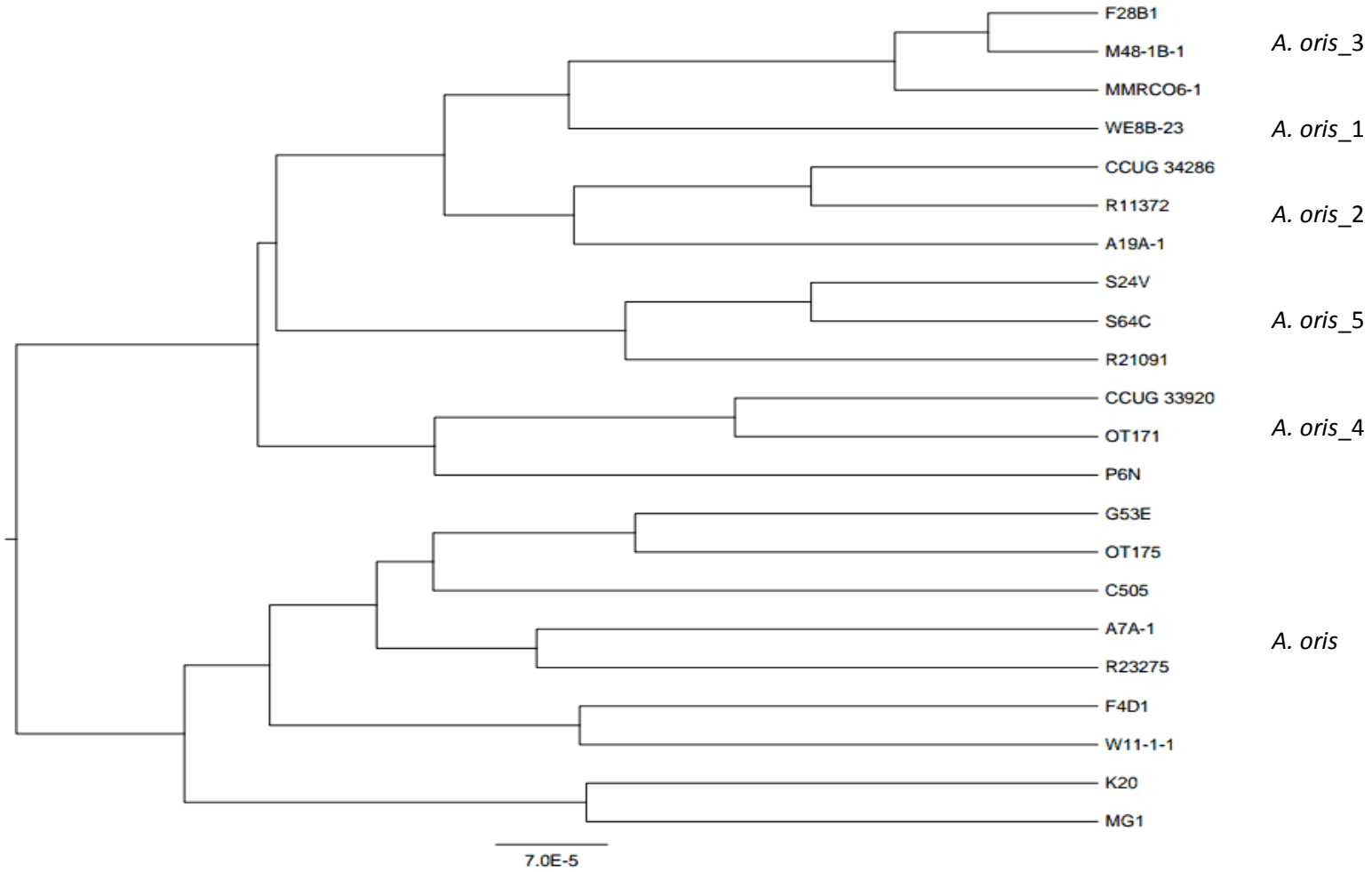


Figure 4.6: ClonalFrame Analysis of *A. oris* isolates.

***Six independent groups were observed known as *A. oris*, *A. oris*_1, *A. oris*_2, *A. oris*_3, *A. oris*_4, *A. oris*_5.**

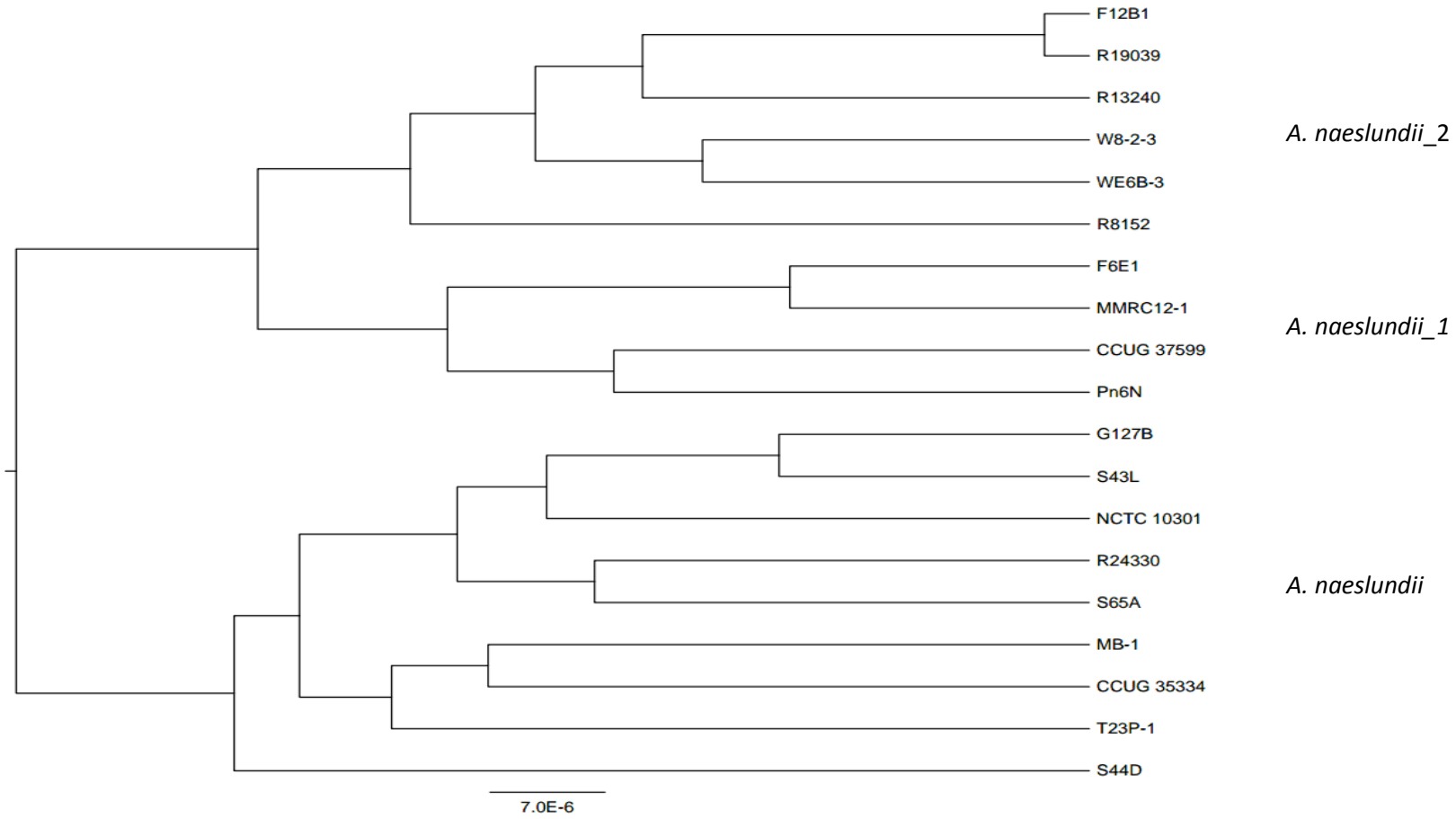


Figure 4.7: ClonalFrame analysis of *A. naeslundii* isolates.

***Three independent groups were observed *A. naeslundii*, *A. naeslundii_1* and *A. naeslundii_2*.**

4.4.2 Core And Pan-Genome Analysis

4.4.2.1 Introduction

In addition to ClonalFrame analysis, Core and Pan Genomes were obtained for each of *A. oris* and *A. naeslundii* separately using PM. Pan-genome of microbial species is defined as the “full complement of genes in a species, and is typically applied to bacteria and archaea, which can have large variations in gene content among closely related strains” (Smokvina *et al.*, 2013; Tettelin *et al.*, 2005). Therefore, “a bacterial species can be described by its pan-genome, which is composed of a "core genome" containing genes present in all strains, and a "dispensable genome" containing genes present in two or more strains and genes unique to single strains” (Medini *et al.*, 2005). The number of genes which are unique is vast therefore pan genome size can be larger than any single genome. The core genome is a set of genes which are orthologous conserved in all species.

4.4.2.2 Material And Method

To make the plots, an output file after a progressive mauve run that ends with ".backbone" is required. The backbone file relates to Locally Collinear Block's present in two or more genomes used to identify the nucleotide coordinates in each LCB. This file contained a row of tables for each genomic region and its presence/absence/location was mentioned in columns in each genome in the respective file. The number of shared or core genes were extracted using in-house programming.

4.4.2.3 Results

Core genome is in red and pan genome is in blue (Figure 4.8 & 4.9). The analysis revealed that the core genes decreased as more sequenced strains were added in the analysis. The maximum core genome size calculated was ≤ 2.7 Mbs for 22 *A. oris* genomes and ≤ 1.5 Mbs using 19 *A. naeslundii* genomes. Similarly the pan genome size calculated was ≥ 9.3 Mbs for 22 *A. oris* genomes and ≥ 5.7 Mbs using 19 *A. naeslundii* genomes. The core genome size of *A. oris* is smaller as compared to the core genome size of *A. naeslundii*, while the pan genome size of *A. oris* is higher as compared to the pan genome size of *A. naeslundii*. As more and more genomes (x-axis) were considered (Figure 4.8 & 4.9), the sum of the length of regions found in all of them (i.e. core in red) decreases whereas the sum of the length of regions found in at least one genome (i.e the pan in blue) increases. This happened faster for *A. oris* than for *A. naeslundii*, presumably because the genomes in *A. oris* are more closely related to each other than the genomes in *A. naeslundii*. This shows the greater diversity of *A. oris* as compared to *A. naeslundii* (Figure 4.8 and 4.9).

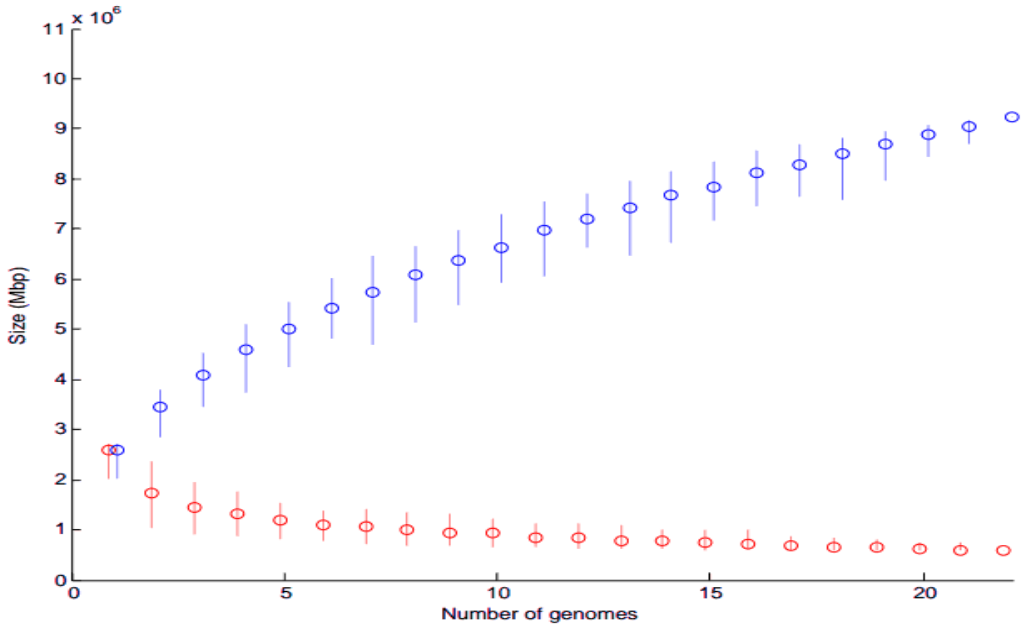


Figure 4.8: Core and Pan-genome model of 22 *A. oris* strains.

*** Core and pan genome size is plotted as a function of a number (n) of genomes added. Circles are the average of such values. A red circle represents core genomes while a blue circle represents pan genomes of *Actinomyces*.**

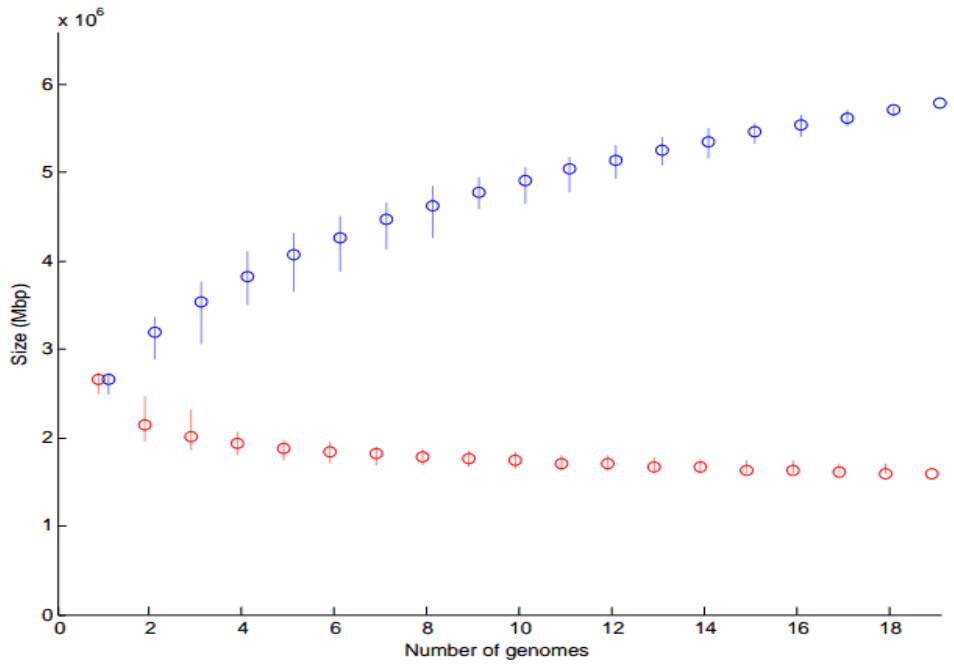


Figure 4.9: Core and Pan-genome model of 19 A. naeslundii strains.

***Core and pan genome size is plotted as a function of a number (n) of genomes added. Circles are the average of such values. A red circle represents core genomes while a blue circle represents pan genomes of *Actinomyces*.**

4.5 Mapping And Alignment Of Core Genomes Of *Actinomyces* Using Gene-to-Gene Analysis

4.5.1 Material And Methods

PM tends not to work with large number of genomes (more than 25), therefore the need to align 43 genomes in one task was essential to quickly compare multiple whole genome sequences which were a million base pairs in length. This comparison facilitates the population study, core genome study and genome evolution study therefore another approach of gene-to-gene alignment was adapted and core and shared genes were identified. Mapping between the genes (proteins) of the 43 annotated genomes delivered in this study was done using RAST (http://rast.nmpdr.org/). 10 genomes were aligned showing only those genes (proteins) with >90% homology with *A. oris*-MG1 at a time in RAST and exported to an excel file. This was done using a batch of 10 genomes (this number was the maximum which could be aligned at one time, always using a mixture of *A. oris* and *A. naeslundii*). The composite file was made in Excel deleting those genes not common to all genomes. RAST server has the ability to analyse and align 10 genomes in 10-15 minutes while annotation was set up overnight.

476 genes were identified as core when all strains were considered in the mapping table and for each of these MUSCLE was used to align the sequences (http://www.ebi.ac.uk/Tools/msa/muscle/). The MUSCLE programme used a nonaligned FASTA file as input and produces an output as an aligned FASTA file. To do each gene alignment manually would take longer therefore a little Matlab script was used to do this. This script reads mapping table as well as the gene sequences of each genome to produce the correct input files for MUSCLE. Each gene-by-gene alignment is then appended into a XMFA file, with "=" signs separating the genes, and this is exactly the format needed as input for ClonalFrame.

4.5.2 Results of Core Genome Analysis

4.5.2.1 Evolutionary Analysis

There were 476 single-copy genes universally present in all 22 *A. oris* and 19 *A. naeslundii* examined *Actinomyces* genomes (Complete list of the number of core genes is in Appendix 4A). The phylogeny of the concatenated sequences of 476 genes was constructed using the ClonalFrame analysis programme (version 1.1) (Figure 4.10). ClonalFrame (Didelot & Falush, 2007) is a statistical software programme and was employed to estimate the ratio of probabilities that a given site is altered through recombination and mutation (r/m ratio) and therefore used to infer the recombination and mutation event introduced in micro-organisms during their evolution and to reveal the origin from their ancestor. The ClonalFrame analysis displayed separate clusters of *A. oris* and *A. naeslundii*. The annotated tree is shown in Figure 4.10. The majority of isolates of *A. oris* derived from a single clonal lineage. Only two strains P6N and CCUG 33920 formed a distinct cluster but were close to *A. oris* and therefore assigned to the *A. oris* group and is proposed to share either an ancient common ancestor or formed a distinct lineage. The genealogy of two species was observed and analysis showed few differences. In the genealogy of *A. oris* two strains named CCUG 33920 and OT171 share a common ancestor with P6N and formed a distinct cluster in *A. oris* group. The same is observed for k20 and MG1 which originates from a single branch and share a mutual ancestor with other isolates of this group, named W11-1-1, F4D1, c505, R23275, A7A-1, OT175 and G53E. Another distinct lineage was found in the *A. oris* group having only three isolates named S24V, S64C and R21091. Strains A19A-1, R11372, CCUG 34286, WE8B-23, MMRCO6-1, M48-1B-1 and F28B1 were found to originate from a single branch of a clade in the *A. oris* group and also formed a distinct cluster in the *A. oris* clade. All 19 strains of *A. naeslundii* developed from a clonal lineage. The clonal genealogy was investigated for *A. naeslundii* isolates and they formed a compact cluster as compared to the *A. oris* clade and were shown to have less diversity as compared to isolates belonged to *A. oris* clade. All 19 isolates of AN were

found to divide into three sub groups under the *A. naeslundii* clade. The isolates which grouped in the main *A. naeslundii* cluster were R24330, S65A, NCTC 10301, G127B, S43L, MB-1, CCUG 35334, T23P-1 and S44D but CCUG 37599, Pn6N, F6E1 and MMRC12-1 formed a distinct branch along with another branch having 6 isolates named F12B1, R19039, R13240, W8-2-3, WE6B-3 and R8152. Therefore from seeing the clonal branch pattern it was observed that *A. oris* isolates were divided into six groups and *A. naeslundii* into three groups. *A. johnsonii* and OT171 were assigned separate positions to *A. oris* and *A. naeslundii* due to their distinct origin in the dendogram as seen in Figure 4.10. The findings from this analysis were similar to the one obtained with MLST, DDH, Mauve and PM.

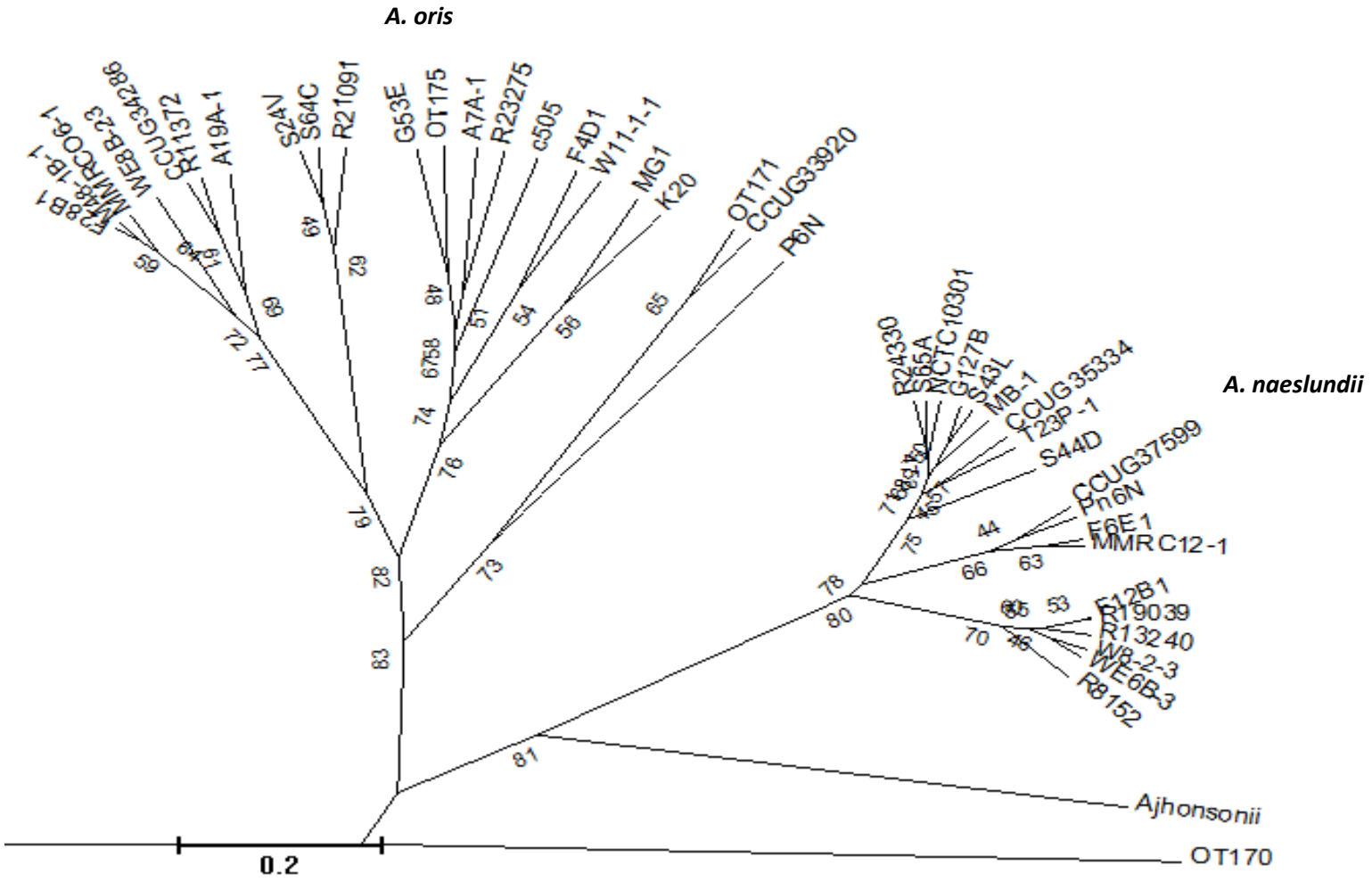


Figure 4.10: ClonalFrame consensus tree for 43 *Actinomyces* genomes using 476 Core genes. Displayed in MEGA 3.1, the numbers are the bootstrap values.

4.5.2.2 Recombination Analysis

The evidence of recombination in each isolate was analyzed using the core genome sequences of each species, *A. naeslundii* and *A. oris,* and for both species the inference of recombination was obtained from ClonalFrame analysis (Figure 4.11). The population structure of the group of *"Actinomyces-naeslundii"* was observed. The overall configuration of the Neighbur-Joining tree is also very similar to that obtained using MLST, Gene-to-Gene Analysis of core genes, Mauve and PM. The number of common genes is low, but perhaps this reflects the diversity of taxa included in the analysis. Each branch (Figure 4.11) of the tree is annotated with two numbers separated by a dash. The first number represents how many mutation events ClonalFrame has detected on that branch. The second number represents how many recombination events ClonalFrame has detected on that branch. Then the relative rates of recombination in the two clades were measured using Excel. The numbers were summed for all the branches in the two clades to get the total number of mutation and recombination events detected in the two clades. Then Chi Square test was applied to compare the relative rates of recombination in the two clades using online tool (http://www.socscistatistics.com/tests/chisquare/). Chi square test was performed and for this a 2x2 contingency table (Table 4.4) was made using the values mentioned in Figure 4.10 of mutation events and recombination events in each category of the two groups, *A. oris* and *A. naeslundii*. The significance level was set to 0.05 to get the chi square values. This contingency table showed observed cell totals, (the expected cell totals), and the chi square statistics for each square. The Chi-square statistic is 21.4866. The P value is $< 4 \times 10^{-6}$. Extensive Genomic Variation within Clonal Complexes of Actinomyces was observed.

Table 4.4: Contingency table for measuring significance of mutation and recombination events in *Actinomyces*.

	A. oris	***A. naeslundii***	**Marginal rows total**
Mutation events	21201 (21057.66) [0.98]	10414 (10557.34) [1.95]	31615
Recombination events	3171 (3314.34) [6.2]	1805 9(1661.66) [12.37]	4976
Marginal columns totals	24372	12219	36591 (Grand total)

* The values before the brackets represent observed cell totals
** The values in brackets showed the expected values of cells in total
*** The values in square brackets represent chi statistics

"rho/theta" and "r/m" are the "two statistics used to assess the relative contribution of recombination and mutation in the creation of the sample from the common ancester". "rho/theta is the ratio of the rates at which recombination and mutation occur. it is therefore a meaure of how often recombination events happen relative to mutations. r/m is the ratio of probabilities that a given site is altered through recombination and mutation. It is therefore a measure of how important the effect of recombination was in the diversification of the sample relative to mutation" (http://www.stats.ox.ac.uk/~didelot/files/clonalframe-userguide.pdf). It is estimated from the ClonalFrame analysis on whole genome conserved genes that rho/theta is equal to 0.22 (with 95% credibility interval [0.19-0.25]) which means that on average recombination is 4 to 5 times less frequent than mutation. On the other hand, r/m (recombination/mutation) was estimated to be equal to 1.34 (interval [1.24;1.62]) which means that recombination introduces slightly more substitutions than mutation does. The way to reconcile these two statements is in the fact that recombination may introduce several nucleotide changes in one go, whereas mutation always only affects a single site. Indeed, the average length of recombined segments was estimated to be equal to 236bp (interval [222; 257]). Analysis revealed that *A. oris* showed more recombination and mutation events than *A. naeslundii* does. In summary it was concluded that *A. oris* and *A. naeslundii* were clearly distinct species with a

characteristic of predominantly clonal structure and genetic stability of core genes, as shown with ClonalFrame analysis, and house keeping genes, as shown using MLST analysis. They apparently originate from a common ancestor, showing closer phylogenetical relationship and recombination events were found to be more prevalent than mutation with the r/m ratio of 1.34. The average recombination rate is higher in *A. oris* as compared to *A. naeslundii* isolates.

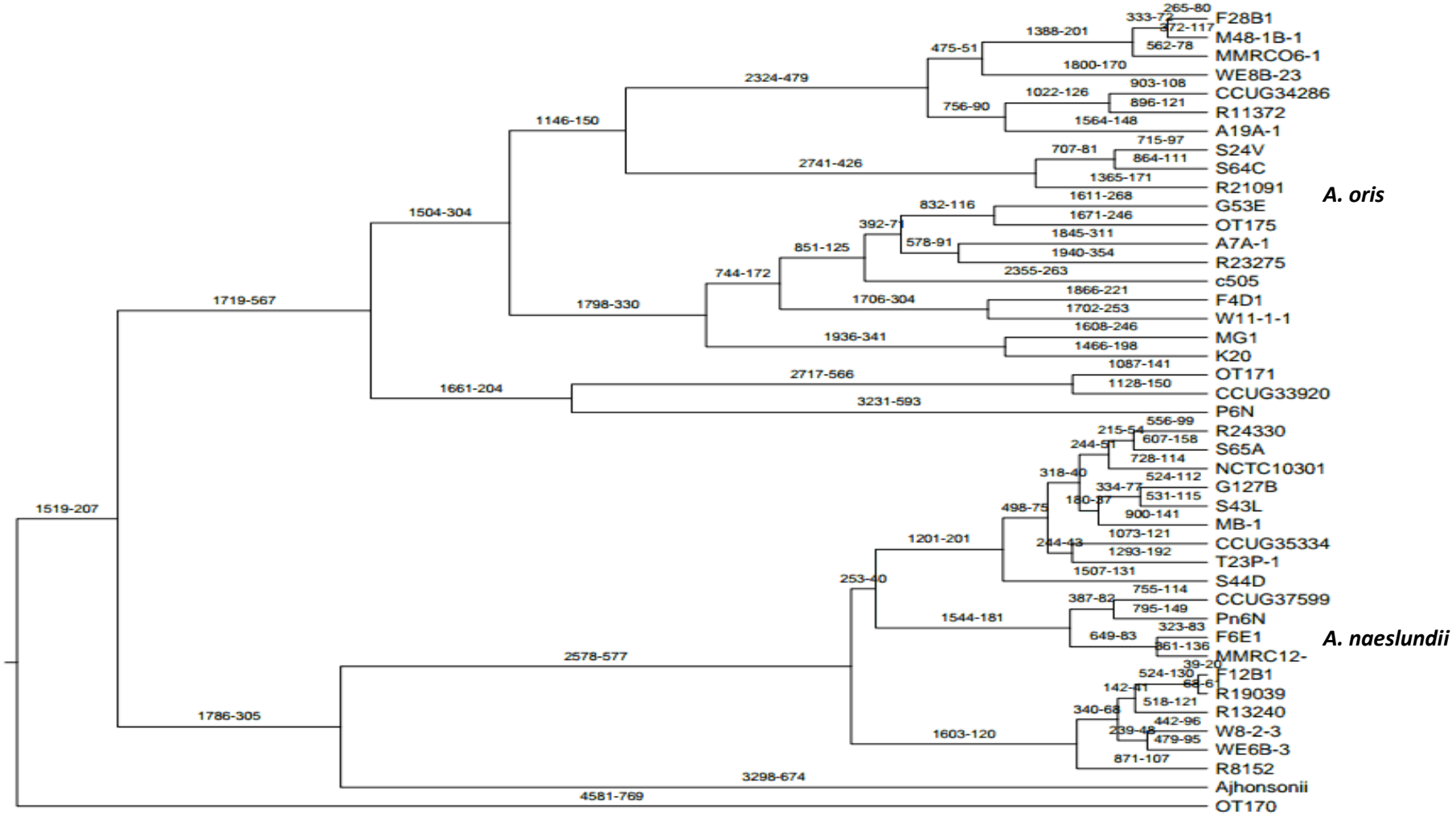


Figure 4.11: Recombination and mutation event locations on each branch of the dendogram showing evolutionary relationships.

4.6 Discussion

Actinomyces are the common bacteria in the oral cavity and comprise a significant proportion of the bacterial flora of dental plaque. As such they are significant players in the complex community of bacteria that exist in this environment and their contribution to the overall balance and bahaviour of the plaque community is likely to be highly significant. *Actinomyces* represents a wide variety of phenotypically and genotypically diverse species. The identification of *Actinomyces* spp. is very difficult and variable (Hall, 1997). Initially *Actinomyces* were identified by their Catalase activity to differentiate among *A. naeslundii* and *A. viscosus* (Ellen & Balcerzak-Raczkowski, 1975; Ellen, 1976; Schaal, 1986). Many studies revealed that *A. naeslundii* and human strains of *A. viscosus* (serotype II, now known as *A. oris*) were highly related and can not be easily differentiated using only phenotypic and serological techniques (Fillery *et al.*, 1978; Gerencser & Slack, 1976; Schofield & Schaal, 1981). A study by Johnson *et al*. (1990) showed that the 51% of stains related to genospecies I (also known as *A. naeslundii*) gave positive catalase activity and also 2% of genospecies II (also known as *A. oris*) showed positive catalase activity.

Effort was also made to distinguish *Actinomyces* species using molecular biology techniques. 16S rRNA gene analysis was the most widely accepted method for species classification. The first big attempt of finding the intragenic relationship revealed that *A. bovis, A. viscosus, A. naeslundii, A. odontolyticus* and *A. israelli* were all genetically distinct species (Stackebrandt & Charfreitag, 1990). Another study shed light on the phylogeny of 22 different *Actinomyces* species isolates, including type strains of *A. naeslundii, A. viscosus* and *Actinomyces* serotype WVA 963 and they found 98.8% sequence similarity between *A. naeslundii* (genospecies I) and animal *A. viscosus* isolates (Ramos *et al.*, 1997). Furthermore, 97.9% homology was observed of *Actinomyces* serotype WVA 963 to *A. naeslundii* and 97.7% was investigated with *A. viscosus* showing that they are not significantly different. Moreover, novel oligonucleotide probes were designed to distinguish *A. naeslundii* and *A. viscosus* in

DNA hybridization assays but only partial success was gained (Tang *et al.*, 2003). The limitation for using the oligonucleotide probes was that the probe which was designed for identifying genospecies I was also detecting catalase negative strains of genospecies II and similarly the probe for *A. viscosus* was also mistaken with detecting catalase positive strains of *A. naeslundii* genospecies II.

Restriction fragment length polymorphism (RFLP) was another approach adopted in an attempt to distinguish *Actinomyces* species. RFLP used restriction enzymes which cut the DNA fragment at specified positions resulting in a number of smaller fragments. Type strains of seven *Actinomyces* species which were commonly found in the human mouth were discriminated properly using 16S DNA-RFLP technique by using restriction enzyme *M*nII while *Hae*III, *Hpa*II, and *Cfo*I enzymes showed the same restriction pattern for distinct species (Sato *et al.*, 1998). This method was successfully employed to classify five clinical isolates of *A. naeslundii* genospecies I and nine isolates of genospecies II. Moreover, in another study variant band pattern was observed among individual strains in both *A. naeslundii* genospecies I and II by using the *M*nII restriction enzyme which shed light on the diversity of the 16S rDNA of each genospecies (Ruby *et al.*, 2002).

Another approach called "amplified 16S ribosomal DNA (rDNA) restriction analysis" (ARDRA) was employed. PCR products of various *Actinomyces* species were treated with restriction enzymes (*Hae*III;*Hpa*II) (Hall *et al.*, 1999). *Hae*III enzyme was able to differentiate two taxa of *Actinomyces* genospecies I and II. However, the DNA profiles observed with ARDRA technique showed less than 99% sequence similarity to the one observed with 16S rDNA sequence of the respective genospecies type strains.

Other techniques which were employed in an attempt to classify *Actinomyces* were AP-PCR, chromosomal DNA fingerprinting, checkerboard DNA-DNA hybridisation and

REP-PCR. Whole genomic DNA was amplified with 40 arbitrary primers (AP-PCR) and the study revealed high sequence heterogeneity among all seven *A. naeslundii* genospecies I and II strains (Ruby *et al.*, 2002). The chromosomal DNA fingerprints gave a high number of bands in the gel with *Sma*I digestion therefore it was very difficult to discriminate among species and also extensive heterogeneity was observed when using 18 mother-child pairs (Ruby *et al.*, 2002). Another study by Barsotti supported the same findings (Barsotti *et al.*, 1993). Checkerboard DNA-DNA hybridization technique was also not very successful as it showed cross reactions between five strains of *A. naeslundii* despite using highly stringent conditions (Ximenez-Fyvie *et al.*, 1999). The limitation of inter-laboratory differences in the results were observed while using checkerboard DNA-DNA hybridization assays. High heterogeneity among *A. naeslundii* genospecies II (currently known as *A. oris*) was observed using repetitive extragenic palindromic PCR (REP-PCR) (Paddick *et al.*, 2003). In this study, 78 genotypes were identified in 246 isolates of *A. naeslundii* genospecies II among seven caries-free subjects and 56 genotypes were found in 208 strains isolated from seven caries-active subjects.

The potential ambiguities were observed while classifying the *Actinomyces* which, indicated the need for new techniques to differentiate between different species and so the technique of housekeeping gene analysis was introduced. This new approach was employed to identify those closely related species which failed to be recognized with 16S rRNA gene analysis (Arif *et al.*, 2008; Christensen *et al.*, 2004; Kiratisin *et al.*, 2005; Poyart *et al.*, 1998; Wertz *et al.*, 2003). In a study the pairwise distance was calculated among seven taxa of enteric bacteria and it was investigated that the distances were higher in five genes named, *gap*A [glyceraldehyde-3-phosphate dehydrogenase], *gro*EL [chaperone-encoding gene], *gyr*A [DNA gyrase A subunit], *omp*A [outer membrane protein II] and *pgi* [glucosephosphate isomerise] as compared to *rrs* gene [16S rRNA] which is very helpful to clearly distinguish bacteria of different taxa (Wertz *et al.*, 2003). In another study three housekeeping genes were studied to distinguish bacteria of the family *Pasteurellaceae*; *atp*D [ATP synthase subunit beta], *infB*

[translation-initiation factor IF2] and *rpoB* [RNA polymerase-beta subunit], while they were not distinguishable with 16S rDNA analysis (Christensen *et al.*, 2004). *A. naeslundii* genospecies I, II, serotype WVA 963 and *A. viscosus* were well distinguished using four housekeeping genes (*atp*A, *met*G, *rpo*B and *gyr*A) and intra-species heterogeneity was observed. Only two genes *atpA* and *metG* are capable of discriminating clearly between *Actinomyces* genospecies 1 and 2 (Henssge, 2009).

The MLST technique used in a study by Henssge (2009) was considered to be the most reliable technique but the limitation of the technique was that it could not differentiate *A. johnsonii* therefore no information was obtained about the sequence diversities within this group of genospecies. In the previous study by Henssge, a strain CCUG 33920 was received as genospecies 2, but resembled four housekeeping genes [*atp*A, *gyr*A, *glt*A, *pgi*] of *A. johnsonii* and was also identical to *met*G and *rpo*B of *A. johnsonii*. In the past the classification of *A. johnsonii* was very controversial, therefore it was regarded as *A. naeslundii* serotype IV based on serological findings (Gerencser & Slack, 1976) but this was not accepted. In a study by Schofield (1981), *A. johnsonii* was placed close to the *A. israelli* cluster and later on it was placed closer to the *A. naeslundii*/*viscosus* cluster. The agglutination data suggested that this genospecies could be assigned into genospecies 2 (Putnins & Bowden, 1993). Johnson, based on DNA homology data, placed *A. johnsonii* separate to genospecies 1 and 2 but taken together his findings, Henssge findings and the digital-DDH results from the present study indicate that *A. johnsonii* serotype WVA 963 can be regarded as a separate species and the name of *A. johnsonii* was kept as suggested by Henssge (2009).

The strain P6N and CCUG 33920 were classified as *A. naeslundii* genospecies 1 previously but is genetically related to genospecies 2. These strains were classified previously based on multivariate statistical analysis of phenotypic characteristics, serological reactions and protein binding patterns of cell extracts as genospecies 1 and they were isolated from tooth surfaces of the same person (Hallberg *et al.*, 1998).

Partial house keeping gene analysis of these strains showed that these species can be placed in genospecies 2.

Therefore, the high variability especially among the group of genospecies 2 (*A. oris*), does not allow the clear indication of differentiation among two genospecies. The availability of complete genome sequences of *Actinomyces* has not only shed light on the genetic features, but also provide the basis for the application of post-genomic techniques. The genome sequence identified several large gene clusters, suggesting they had been acquired by horizontal gene transfer. However, the sequence of the genomes revealed few mechanisms by which genetic diversity can be generated.

4.6.1 MLST Analysis

A detailed previous taxonomic study using 6 concatenated genes sequenced by Henssge (2009) delineated distinct clusters for species of *Actinomyces*. However using 7 concatenated genes some strains of each species were quite distinct and formed separate clusters. To determine if these clusters represent distinct subspecies, studies have been undertaken to obtain whole genome sequences of these isolates. In section 4.1 of MLST analysis seven more publically available strains were added in the previous study to investigate the phylogenetic status (Figure 4.1). A MLST analysis approach was used to investigate the population diversity and structure and therefore this analysis sheds light on the bacterial genomic evolutionary process, classification and ecology. In epidemiological studies, MLST is very useful where the origin of pathogens and its spread is the main concern. To date, MLST analysis has been applied in many research areas. *Neisseria* spp. (Bennett *et al.*, 2007; Birtles *et al.*, 2005; Kriz *et al.*, 2002; Maiden *et al.*, 1998; Maiden, 2006), *Streptococcus* spp. (Enright & Spratt, 1998; Jones *et al.*, 2003; King *et al.*, 2002; Pullinger *et al.*, 2006), *Campylobacter* spp. (Dingle *et al.*, 2001; Miller *et al.*, 2005; Sails *et al.*, 2003; Schouls *et al.*, 2003) and *Staphylococcus* spp. (Enright *et al.*, 2000; Kozitskaya *et al.*, 2005; Robinson & Enright, 2004) were the most

often investigated species. Six to ten loci were considered acceptable for multi-locus sequence typing analysis (Maiden, 2006) therefore in the present study seven genes (*atp*A, *glt*A, *gyr*A, *met*G, *pgi*, *phe*S and *rpo*B) were selected. According to Neighbor-Joining tree analysis, *A. naeslundii* isolates were arranged in one single cluster using 7 housekeeping genes while few splits were observed in *A. oris* which illustrates the diversity of *A. oris* isolates.

4.6.1.1 Enhanced Species Classification

The phylogenetic trees of concatenated sequences of *A. oris* and *A. naeslundii* (Figure 4.1) revealed 14 isolates as distinct to *A. oris* clonal complex which shows that these isolates are from 14 different sources/individuals. The isolate with study number OT171 was found in the subcluster comprising three isolates and isolate MG1, k20 and c505 were found in the main *A. oris* cluster, therefore suggesting that OT171 is a subspecies of *A. oris*. The same was observed in a study by Henssge where strain number CCUG 34286 was found in a separate cluster of seven isolates. This isolate was previously identified as *A. naeslundii* serotype III (Johnson *et al.*, 1990) and emended within a group of *A. oris* (Henssge *et al.*, 2011) as its subspecies. In conclusion, this study supports the hypothesis that there are perhaps subspecies of *A. oris*.

There was found to be no recombination event within or between the species using house keeping genes analysis (Henssge, 2009).

4.6.2 Digital DNA-DNA Hybridization Analysis

The interesting feature of the present study results is that the taxonomy obtained with most advanced techniques was in agreement with the classification investigated by Johnson (1990). Therefore the current homology data and the findings reported by Johnsons suggested that DNA homology is the most reliable method either by wet-

laboratory or in-silico methods but the advantage of in-silico methods over wet-laboratory methods is that it is less tedious and error prone and represents the accurate measurement of DNA homology values leaving any ambiguities of variations of data obtained with wet-laboratory techniques. It is proposed that *A. naeslundii* should be kept as single cluster and *A. oris* is divided into 6 clusters as suggested previously by Henssge (2009). The study described in section 4.2 represented the first Digital-DDH investigation of *Actinomyces* species and required the availability of finished whole genomes. In addition to determining which species are related to which group, it provides researchers with more accurate taxonomical classification at species level. Mean DDH values were determined between 43 genome sequenced strains belonging to six phylogenetically distinct groups of *A. oris* and three of *A. naeslundii*. Strains within a group were closely related, i.e. showing > 60% homology values and they had > 94% 16S rRNA gene sequence identity (Henssge, 2009) for *A. oris*. The six *A. oris* groups from AO to AO_5 have more than 17 isolates sequenced including the internet strains. The names of these isolates were mentioned in Table 2.1. This group represents strains from caries free and caries active individuals. The isolates were only having Gram-positive bacteria. They comprise a consistent genome size from 2.9MB (W11-1-1) to mean genome sizes of 3.5Mb (WE8B-23), which allowed for robust interpretations.

The microplate DNA-DNA hybridization method of Ezaki *et al*., is a well established and frequently used method in bacterial taxonomy previously (Ezaki *et al.,* 1989). Several reviews (Rossello-Mora & Amann, 2001; Stackebrandt & Liesack, 1993; Stackebrandt & Goebel, 1994; Stackebrandt & Rossello-Mora, 2006) cite that 80% identity must be shared in DNA fragments in order to hybridize during DDH experiments while the current results reveal that the 70% DDH recommendation covers relatively identical strains at the genomic level, and previous studies results based on the phenotypic similarity of strains were reliable using the same 70% DDH standard. The authenticity of the values obtained using in-silico DDH comparisons emphasize the call for the concept of species to be reevaluated and it was discussed in a committee and recommendations were: “Investigators are encouraged to propose new species based

upon other genomic methods or techniques provided that they can demonstrate that, within the taxa studies, there is a sufficient degree of congruence between the technique used and DNA-DNA reassociation. In addition, investigators are encouraged to develop new methods to supplement or supplant DNA-DNA reassociation" (Stackebrandt *et al.*, 2002). This recommendation was met in the present study by employing the Digital DNA-DNA Hybridization technique to obtain the homology values which were authentic and accurate by using advanced whole genome sequences. Separate species were considered to have difference of up to 21 % in gene content between strains, e.g. up to 1000 genes may be different in a genome of approximately 5Mb size (Goris *et al.*, 2007).

The NCBI-BLASTN was considered the best performing programme to calculate distances among genomes for digital-DDH calculations. NCBI-BLASTN is comparatively fast and took a feasible amount of RAM only. The whole genome of selected strains of *Actinomyces* were sequenced for the first time in this study but the strains have been used in various studies since 1990 (Do *et al.*, 2008; Henssge *et al.*, 2009; Henssge *et al.*, 2011; Johnson *et al.*, 1990). Whole genome analysis, as well as core genome and pan genome analysis of 43 whole genome sequenced strains revealed that *Actinomyces* genomes are highly heterogeneous. These findings are consistent with the previous studies using MLST analysis and DNA homology analysis using wet-laboratory techniques.

4.6.3 ClonalFrame Analysis

ClonalFrame analysis was employed in the current study to explore the evolutionary history, recombination and mutation rate. A Neighbor-Joining tree were obtained after ClonalFrame analysis using MEGA version 3.1. The tree revealed long separate branches for isolates which indicates the presence of few isolates on one cluster. The tree like structure clearly indicates that *A. naeslundii* was distinct from the majority of *A.*

oris except for a few isolates which were found to be present in the intermediate position and there was an indication of some branches within each species. ClonalFrame analysis revealed that *A. oris* and *A. naeslundii* originates separately when a combined analysis was obtained on 43 genomes. There were clonal lineages observed in both species with few splits due to recombination and mutation events (Figure 4.10 and 4.11). The separate analysis revealed a hierarchical structure with high resolution for *A. oris* as compared to *A. naeslundii* which showed a big compact split for *A. naeslundii* that represents one major cluster for the population and two separate independent clusters in *A. naesludnii* while five separate clades were found for *A. oris* with one main cluster (Figure 4.10 and 4.11)**.** The similar structure of two lineages was observed for *Vibrio vulnificus* (Bisharat *et al.*, 2007). It was observed that strains causing human diseases were more clonally arranged due to frequent recombination and mutation while environmental strains were less prone to recombine and mutation as shown with less resolution due to smaller recombination rates among them. In conclusion the availability of whole genome sequences of these isolates provide insight into the intricate genomic features of *Actinomyces*, and will surely serve as an important reference for those interested in the study of oral *Actinomyces* strains associated with plaque in the future. Overall there seems to be some patterns that indicate evolutionary events in the history of *Actinomyces as* species. Furthermore there is a large ancestral split between *A. oris* and *A. naeslundii* strains. The evolution of two big clades among the *Actinomyces* spp. suggested that these strains evolved from a single ancestor many million years ago and evolved into two main groups which had homologous genes.

4.6.4 Recombination Analysis

Recombination events (RE) were observed on core genomes of isolates (43 isolates of *Actinomyces* spp.) and RE were observed in all *A. oris* isolates as well as in 19 *A. naeslundii* isolates. Previously no significant RE were observed in housekeeping genes while some RE were observed between *A. oris* and *A. naeslundii nan*H gene which codes for sialidase. RE were also observed in only two isolates of *A. oris* (S62B and

S64C) and three isolates of *A. naeslundii* (G51C, S41C and CCUG 34725; (Do *et al.*, 2008). The sequencing results gave longer than 400bp DNA fragments. Sialidase was not a housekeeping gene and has the ability to desialylate IgA (Frandsen, 1994) and used for the purpose of nutrient gaining (Beighton *et al.*, 1986; Lucas *et al.*, 1997; Smith & Beighton, 1986).

The levels of recombination vary in bacteria in housekeeping genes. The Recombination rates was observed in bacteria in a study and it was investigated that recombination varies greatly as compared to point mutation (r/m) (Vos & Didelot, 2009). There are the species such as *Flavobacterium psychrofilum*, *Vibrio parahaemolyticus, Salmonella enterica, Helicobacter pylori* and *S. pneumoniae* showed very high recombination rates and the r/m ratio observed was between 13.6 and 63.6 while *Staphylococcus aureus, P. gingivalis, L. casei, O. oeni* and *E. faecalis* had less than 1 r/m ratio which shed light on the changes in nucleotide sequences being due to point mutation rather than recombination. Recombination can also be explained with the association levels (I_A). The I_A in *A. oris* and *A. naeslundii* are 0.21 and 0.231 respectively and significant linkage disequilibrium was observed in both species. The I_A is 0.04 for those species who recombine frequently, as was observed for *N. gonorrhoeae* (Smith *et al.*, 1993) and *P. gingivalis* showed 0.09 (Koehler *et al.*, 2003) and 0.047 I_A was observed for *Microcoleus chtonoplastes* (Lodders *et al.*, 2005).

4.6.5 Core Gene Analysis

The present study provides the first insight and novel approach for defining the core genome of an *A. oris* and *A. naeslundii* species for taxonomical classification. The core genome was estimated from the whole genome sequences of 43 *Actinomyces* strains as 476 orthologs or gene families. The function of these gene families is given using RAST annotations or subsystem technology but still the functions of all gene families is not known and they were characterized as hypothetical proteins. This finding indicates that there is still need to learn about the functions of the highly conserved and essential genes in *Actinomyces* species. The phylogenetic tree based on core gene sequence

similarity is the same as the phylogenomic tree based on total genome content (Figure 4.5). There are the sets of strains clustered together in similar mode using full genomic content and core genes (ClonalFrame analysis) dendograms suggesting that they are closely related in evolution. In the present study the annotation of genes is in agreement with the observed phenotype e.g for lacto-*N*-biose utilization pathway. Additional experiments such as knocking out target genes of Lacto-*N*-Biose Phosphorylase were undertaken have increased the understanding of the type of relation between the target and its phenotype (Knock-out gene experiments are explained in Chapter 5). This observation supports the facts that understanding of more complex phenotypes could be observed especially where the link was less clear between matching genes and its phenotype. The predicted genes can be used for the large scale screening of particular strains from the culture collections easily. Finally the pan genome analysis also shed light on the variability of cell surface proteins and exopolysaccharides. The descending trend in the core genome size was also observed in many studies with the increasing number of isolates (Chen *et al.*, 2013)

4.6.5.1 Conclusion And Practical Implication Of Dispensable Gene Study

The availability of multiple genome sequences is crucial to expanding the existing knowledge of species concept. In a recent study it has been investigated from the knowledge of dispensable genes that the design of a universal vaccine was made possible against group B *Streptococcus* (Maione *et al.*, 2005). This is a clear practical demonstration of a use of sequences obtained from multiple genomes of a single strain which is pathogenic for the design of vaccine against that pathogen.

Taken all the data together from MLST, Digital DDH, Mauve, Core and Pan genome analysis, it can be said that *A. oris* and *A. naeslundii* are distinct species. Their clonal structure was characterised and the results revealed that both species originate from a common ancester. This finding showed the close phylogenetic relationship between the two genospecies. The recombination and mutation events were characterised in both species showing that they were linked together evolutionarily.

Chapter 5 CHARACTERIZATION OF "Lacto-N-Biose" OPERON
&
LNB UTILIZATION BY *Actinomyces oris*

5.0 Introduction

5.0.1 Characterisation Of Lacto-*N*-Biose Operon

The ability of bacteria to grow and proliferate in the mouth is determined in large part by their ability to obtain nutrient from the local environment, especially from mucins and salivary glycoproteins, since dietry foods are rapidly removed from the mouth (Keene *et al.*, 1996) and provide novel nutrients only briefly. Mucins possess O-linked glycans, which may form up to 80% of their mass, while glycoproteins contain N-linked glycans. The O- and N-linked glycans are degraded by a variety of exo- and endo-glycosidases produced by the microbiota of the oral biofilm (Klein *et al.*, 1992). The growth of bacteria in the oral biofilm may be a cooperative effort in which bacteria interact synergistically to degrade glycans and liberate sugars, including sialic acid, galactose, N-acetylglucosamine, N-galactosamine, mannose and fucose for growth (Beighton *et al.*, 1986; Homer & Beighton, 1992).

The disaccharides lacto-*N*-biose and galacto-*N*-biose may also be liberated from N- and O-linked glycans respectively. Little is known about the ability of oral bacteria to utilize these sugars although their metabolism has been extensively studied in relation to the influence of human breast milk on the infant gut flora and a novel pathway for the utilization of lacto-*N*-bioside and galacto-*N*-bioside has been described in *Bifidobacterium longum* NCC2705 (Kitaoka *et al.*, 2005). The gene cluster coding for the utilization of lacto-/galacto-*N*-biose consists of three genes coding for components of an ATP-binding cassette-type sugar transporter (BL1638-BL1640), a lacto/galacto-*N*-biose phosphorylase (BL1641), an N-acetylhexosamine phosphorylase (BL1642) and a galactose-1-P uridylyltransferase (Fujita *et al.*, 2005; Yoshida *et al.*, 2012). The presence of this pathway enables *B. longum* and other bifidobacteria, including *B. bifidium*, *B. breve* and *B. animalis* to utilize lacto-*N*-biose for growth while other species were either unable to use it for growth of only minimally. Central to the utilization of lacto-/galacto-*N*-biose is the lacto-/galacto-*N*-biose phosphorylase which catalyses the phosphorolysis of either lacto-*N*-biose or galacto-*N*-biose to form N-acetylglucosamine or N-acetylgalactosamine, respectively and galactose-1-P. The ability of bifidobacteria

to utilize LNB for growth is believed to be the reason this species proliferates in the digestive tract of breast fed infants. It may therefore be expected that oral bacteria with this ability would have an advantage in the oral cavity where LNB could be derived from the O- and N-linked glycans of salivary components. In this study we report the presence of a LNB operon in *A. oris* and the ability of this species to utilize LNB for growth while in the closely related species, *A. naeslundii*, this operon is dysfunctional and subject to extensive degradation.

5.0.2 Arrangement of Genes In And Around LNB Operon In *A. oris*-MG1

The arrangement of genes in and around of LNB operon is highlighted in *A. oris*-MG1 in gene location diagramm as shown in Figure 5.1.

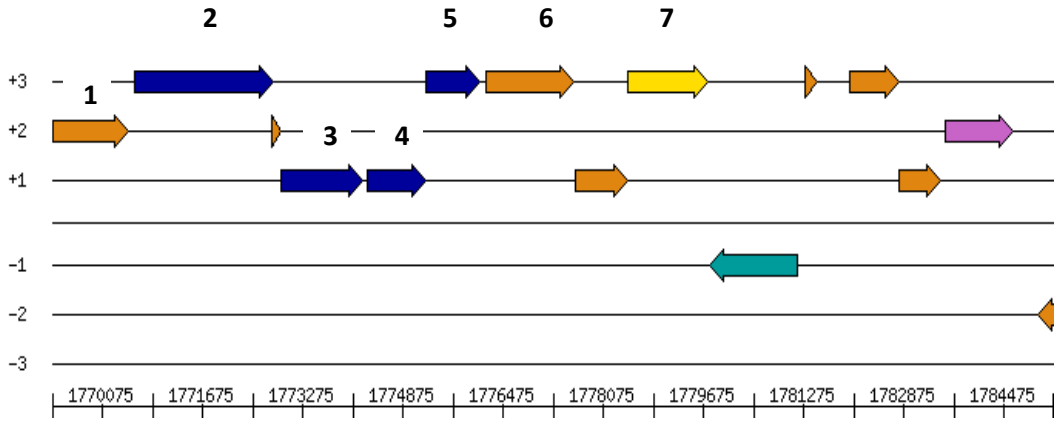


Figure 5.1: *A. oris*-MG1 genome highlighting the gene locations of Lacto-N-Biose Operon.

1. Transcriptional Regulator, RoK family protein (1st orange bar)
2. Lactose-N-Biose Phosphorylase (long blue)
3. Predicted galacto-N-Biose-/lacto-N-Biose (ABC Trasporter, substrate binding protein (2nd blue)
4. Predicted galacto-N-Biose-/lacto-N-Biose (ABC Trasporter, permease component I (3rd blue)
5. Predicted galacto-N-Biose-/lacto-N-Biose (ABC Trasporter, permease component II (4th blue)
6. GlcNAc Phosphomutase (next to blue)
7. Glucose-1-Phosphate thymidyl transferase (next)

5.1 Preliminary Gene Differences In LNB Operon

The whole genome sequence data of isolates were observed after annotation in RAST. Differences were found between *A. oris* and *A. naeslundii* strains. The Lacto-*N*-Biose operon was apparently absent from *A. naeslundii* but present in *A. oris* isolates. These genes are Lacto-*N*-biose phosphorylase (ANA_1637), Predicted galacto-*N*-biose/lacto-*N*-biose I ABC transporter, substrate-binding protein (ANA_1638), Predicted galacto-*N*-biose-/lacto-*N*-biose I ABC transporter permease component 1 (ANA_1639) and Predicted galacto-*N*-biose-/lacto-*N*-biose I ABC transporter permease component 2 (ANA_1640). These genes were found in all *A. oris* sequences in databases (in all 17 strains) but in none of the 19 *A. naeslunsdii* genome sequences. These genes were searched for in the CMR and RAST websites and they have different names as shown in Table 5.1.

Table 5.1: The names of Lacto-*N*-biose gene cluster

Gene number	RAST	CMR
***ANA_1633 (**ANA_1450)**	ABC-type proline/glycine betaine transport systems	glycine betaine transport system permease protein (proX)
***ANA_1634 (**ANA_1451)**	L-proline glycine betaine ABC transport system permease protein ProW (TC 3.A.1.12.1)	glycine betaine transport system permease protein (permease)
***ANA_1635 (**ANA_1452)**	L-proline glycine betaine ABC transport system permease protein ProV (TC 3.A.1.12.1)	glycine betaine ABC transporter ATP-binding protein [3.6.3.32]
***ANA_1636 (**ANA_1453)**	Xylose repressor XyIR (ROK family)	putative xylose repressor
***ANA_1637 (**ANA_1454)**	Lacto-*N*-biose phosphorylase (EC 2.4.1.211) {1,3-beta-galactosyl-N-acetylhexosamine phosphorylase}	conserved hypothetical protein
***ANA_1638 (**ANA_1456)**	Predicted galacto-*N*-biose-/lacto-*N*-biose, ABC transporter, substrate-binding protein	solute binding protein of ABC transporter for sugars
***ANA_1639 (**ANA_1457)**	Predicted galacto-*N*-biose-/lacto-*N*-biose, ABC transporter, permease component 1	permease of ABC transporter for sugars (permease)
***ANA_1640 (**ANA_1458)**	Predicted galacto-*N*-biose-/lacto-*N*-biose, ABC transporter, permease component 2	permease of ABC transporter for sugars (permease)
***ANA_1641 (**ANA_1459)**	GlcNAc phosphomutase (EC 5.4.2.3)	Phosphomannomutase (pmm) [5.4.2.8]
***ANA_1642 (**ANA_1460)**	glucose-1-phosphate thymidylyltransferase (EC:2.7.7.24)	Glucose-1-phosphate thymidylyltransferase (grad)
***ANA_1644 (**ANA_1461)**	Galactokinase (EC 2.7.1.6)	N-acetylgalactosamine kinase (GalNAc kinase) (Galactokinase 2), putative [2.7.1.-]

*** ANA_1633 to ANA_1644 are the gene numbers of *A. oris*-MG1 in CMR website**

****ANA_1450 to ANA_1461 are the gene numbers of *A. oris*-MG1 in RAST website**

5.2 The Subsystem Lacto-*N*-Biose I And Galacto-*N*-Biose Metabolic Pathway Comparison Of *Actinomyces* Spp. With *B. longum* In RAST

The subsystem Lacto-*N*-Biose I and Galacto-*N*-Biose metabolic pathway was observed in sequenced *A. oris* and *A. naeslundii* isolates using the RAST programme and was viewed in SEED under organism tool (Figure 5.2 & 5.3) and the requested subsystem was displayed in subsystem spreadsheet and compared with the *Bifidobacterium longum* LNB pathway which was presented in detail in published papers (Kitaoka *et al.*, 2005; Nishimoto & Kitaoka, 2007b; Wada *et al.*, 2007). All sequenced *A. oris* strains have Lacto-*N*-Biose phosphorylase (EC 2.4.1.211) LnpA), Predicted galacto-*N*-Biose-/lacto-*N*-Biose I ABC transporter periplasmic substrate binding protein (ABCsb) , Predicted galacto-*N*-Biose-/lacto-N-Biose I ABC transporter permease component I (ABCp1), Predicted galacto-*N*-Biose-/lacto-*N*-Biose I ABC transporter permease component 2 (ABCp2) while *B. longum* has Lacto-*N*-Biose phosphorylase (EC 2.4.1.211) (LnpA), N-acetylhexosamine 1-kinase (LnpB), UDP-glucose hexose 1-phosphate uridylyltransferase (LnpC), UDP-glucose 4-epimerase (LnpD), Predicted galacto-*N*-Biose-/lacto-*N*-Biose I ABC transporter periplasmic substrate binding protein (ABCsb), Predicted galacto-*N*-Biose-/lacto-*N*-Biose I ABC transporter permease component I (ABCp1) and Predicted galacto-*N*-Biose-/lacto-*N*-Biose I ABC transporter permease component 2 (ABCp2) gene cluster. All of the 19 sequenced *A. naeslundii* strains were subjected to analysis to see the lacto-*N*-biose operon. None of the genes of LNB operon appeared in *A. naeslundii* isolates (Figure 5.3). Only five of *A. naeslundii* appeared in the comparative analysis with *Bifidobacterium longum* and these were F6E1, CCUG 37599, R24330, MMRC12-1 and Pn6N. They only have degraded LnpD gene as shown in Figure 5.3.

Organism	Domain	Variant[?]	Active	LnpA	LnpB	LnpC	LnpD	ABCsb	ABCp1	ABCp2
		=	yes							
Actinomyces oris S64C (6666666.15742)	Bacteria	1.x	Yes	2287			1172,418, 421	2286	2285	2284
Actinomyces oris F4D1 (6666666.15720)	Bacteria	1.x	Yes	557			1058,2141,2144	559	560	561
Actinomyces oris R11372 (6666666.15734)	Bacteria	1.x	Yes	731			1347,1376,1379	729	728	727
Actinomyces oris 171(new) (6666666.14766)	Bacteria	1.x	Yes	2398			151, 154, 987	2396	2395	2394
Actinomyces oris R21091 (6666666.15737)	Bacteria	1.x	Yes	159			1459,1462,1490	158	157	156
Actinomyces oris F28B1 (6666666.15723)	Bacteria	1.x	Yes	538			495, 498, 528	540	541	542
Actinomyces oris R23275 (6666666.15738)	Bacteria	1.x	Yes	1458			2123,502, 505	1457	1456	1455
Actinomyces oris S24V (6666666.15740)	Bacteria	1.x	Yes	1748,1753			2098,2101,2168	1749	1750	1751
Actinomyces oris C505 (new) (6666666.14679)	Bacteria	1.x	Yes	534			448, 451, 477	536	537	538
Actinomyces oris CCUG 33920 (6666666.15716)	Bacteria	1.x	Yes	1309,1314			1851,1854,368	1311	1312	1313
Actinomyces oris G53E (6666666.15724)	Bacteria	1.x	Yes	1556			462, 488, 491	1555	1554	1553
Actinomyces oris W11-1-1 (6666666.15748)	Bacteria	1.x	Yes	610			786, 946, 949	612	613	614
Actinomyces oris M48-1B-1 (6666666.15726)	Bacteria	1.x	Yes	901, 902			1169,1172,1202	899	898	897
Actinomyces oris K20 (new) (6666666.14767)	Bacteria	1.x	Yes	1039			1099,1127,1130	1038	1037	1036
Actinomyces oris A19A-1 (6666666.15715)	Bacteria	1.x	Yes	1937			2298,298, 301	1935	1934	1933
Actinomyces oris MMRCO6-1 (6666666.15729)	Bacteria	1.x	Yes	211, 212			695, 725, 728	210	209	208
Actinomyces oris WE8B-23 (6666666.15751)	Bacteria	1.x	Yes	2043			1440,1443,1753	2044	2045	2046
Actinomyces oris OT175 (new) (6666666.14768)	Bacteria	1.x	Yes	658			574, 577, 604	659	660	661
A oris P6N (6666666.15731)	Bacteria	1.x	Yes	614			1763,1789,1792	613	612	611
Actinomyces oris CCUG 34286 (6666666.15717)	Bacteria	1.x	Yes	950			114, 143, 146	949	948	947
Actinomyces oris A7A-1 (6666666.15713)	Bacteria	1.x	Yes	691			743, 769, 772	689	688	687
Actinomyces oris MG1 (new) (6666666.14759)	Bacteria	1.x	Yes	1454			1369,1372,1401	1456	1457	1458
Eubacterium rectale ATCC 33656 (515619.6)	Bacteria	*2.0	Yes	125			1689,2816			
Bryantella formatexigens DSM 14469 (478749.5)	Bacteria	*2.0	Yes	554	178, 2896		2485,2589,4288			
Propionibacterium acnes KPA171202 (267747.1)	Bacteria	1.x	Yes	78				2236	2235	2234
Streptobacillus moniliformis DSM 12112 (519441.4)	Bacteria	2.0	Yes	394			446			
Clostridium perfringens ATCC 13124 (195103.9)	Bacteria	2.0	Yes	530			263, 465			
Clostridium perfringens str. 13 (195102.1)	Bacteria	2.0	Yes	636	1263		349, 572			
Abiotrophia defectiva ATCC 49176 (592010.4)	Bacteria	*2.0	Yes	3045	3046		3049			
Clostridium phytofermentans ISDg (357809.4)	Bacteria	2.0	Yes	1898, 2996,573	566		1153,3464,567			
Bifidobacterium longum NCC2705 (206672.1)	Bacteria	1.0	Yes	1559	1560	1561	1562, 1589	1556	1557	1558
Bifidobacterium longum DJO10A (205913.1)	Bacteria	1.0	Yes	1684	1683	1682	1681	1687	1686	1685
Catonella morbi ATCC 51271 (592026.3)	Bacteria	*2.0	Yes	1496	1005		241			

Figure 5.2: Lacto-N-Biose gene cluster comparison in RAST among Actinomyces oris spp and Bifidobacterium longum NCC2705.

LnpA (Lacto-N-Biose phosphorylase (EC 2.4.1.211), LnpB (N-acetylhexosamine 1-kinase), LnpC (UDP-glucose hexose 1-phosphate uridylyltransferase), LnpD (UDP-glucose 4-epimerase), ABCsb (Predicted galacto-NBiose-/lacto-N-Biose | ABC transporter periplasmic substrate binding protein), ABCp1 (Predicted galacto-N-Biose-/lacto-N-Biose | ABC transporter permease component 1), ABCp2 (Predicted galacto-N-Biose-/lacto-N-Biose | ABC transporter permease component 2). LnpA, ABCsb, ABCp1 and ABCp2 corresponds to gene numbers Ana_1637 to Ana_1640 in *A. oris*-MG1 respectively.

Organism	Domain	Variant[?]	active	LnpA	LnpB	LnpC	LnpD	ABCsb	ABCp1	ABCp2
		=	yes							
Actinomyces naeslundii F6E1 (6666666.15721)	Bacteria	2.0	Yes	91			1201,1204,32			
Actinomyces naeslundii CCUG 37599 (6666666.15719)	Bacteria	2.0	Yes	1804			1735,1738,1763			
Actinomyces naeslundii R24330 (6666666.15739)	Bacteria	2.0	Yes	1944			1509,1535,1538			
Actinomyces naeslundii MMRC12-1 (6666666.15728)	Bacteria	2.0	Yes	678			108, 80, 83			
Actinomyces naeslundii Pn6N (6666666.15732)	Bacteria	2.0	Yes	89			298, 30, 301			
Eubacterium rectale ATCC 33656 (515619.6)	Bacteria	*2.0	Yes	125			1689,2816			
Bryantella formatexigens DSM 14469 (478749.5)	Bacteria	*2.0	Yes	554	178, 2896		2485,2589,4288			
Propionibacterium acnes KPA171202 (267747.1)	Bacteria	1.x	Yes	78				2236	2235	2234
Streptobacillus moniliformis DSM 12112 (519441.4)	Bacteria	2.0	Yes	394			446			
Clostridium perfringens ATCC 13124 (195103.9)	Bacteria	2.0	Yes	530			263, 465			
Clostridium perfringens str. 13 (195102.1)	Bacteria	2.0	Yes	636	1263		349, 572			
Abiotrophia defectiva ATCC 49176 (592010.4)	Bacteria	*2.0	Yes	3045	3046		3049			
Clostridium phytofermentans ISDg (357809.4)	Bacteria	2.0	Yes	1898,2996,573	566		1153,3464,567			
Bifidobacterium longum NCC2705 (206672.1)	Bacteria	1.0	Yes	1559	1560	1561	1562,1589	1556	1557	1558
Bifidobacterium longum DJO10A (205913.1)	Bacteria	1.0	Yes	1684	1683	1682	1681	1687	1686	1685
Catonella morbi ATCC 51271 (592026.3)	Bacteria	*2.0	Yes	1496	1005		241			
Vibrio vulnificus CMCP6 (216895.1)	Bacteria	2.0	Yes	3969			1236,1616,3973,734			
Vibrio vulnificus YJ016 (196600.1)	Bacteria	2.0	Yes	4941			2707,3092,4945			
Bifidobacterium longum subsp. infantis ATCC 15697 (391904.3)	Bacteria	1.x	Yes	2180		2178	2177,524			
Anaerococcus prevoti prevotii DSM 20548 (525919.4)	Bacteria	2.0	Yes	1357			1360			

Figure 5.3: Lactose-N-Biose gene cluster comparison in RAST among Actinomyces naeslundii spp and Bifidobacterium longum NCC2705.

*****LnpA (Lacto-N-Biose phosphorylase (EC 2.4.1.211), LnpB (N-acetylhexosamine 1-kinase), LnpC (UDP-glucose hexose 1-phosphate uridylyltransferase), LnpD (UDP-glucose 4-epimerase), ABCsb (Predicted galacto-NBiose-/lacto-N-Biose ⎪ ABC transporter periplasmic substrate binding protein), ABCp1 (Predicted galacto-N-Biose-/lacto-N-Biose ⎪ABC transporter permease component 1), ABCp2 (Predicted galacto-N-Biose-/lacto-N-Biose ⎪ABC transporter permease component 2). LnpA, ABCsb, ABCp1 and ABCp2 correspond to gene numbers Ana_1637 to Ana_1640 in *A. oris*-MG1 respectively.

5.3 The Comparison of *Actinomyces* spp. Amino Acid Sequence Length With *A. oris*-MG1

The amino acid sequence homology to *A. oris*-MG1 was obtained using RAST programme and annotated genome was viewed or browsed using the SEED viewer (version 2.0). The sequence based comparison was used to line up 10 query genomes using *A. oris*-MG1 as a reference genome. The amino acid sequence length and its identity were observed for each query genome and the gene regions selected were between Ana_1634 to Ana_1644. The results obtained with *A. oris* and *A. naeslundii* were compiled in Table 5.2. All of the *A. oris* isolates showed the presence of complete gene length for Ana_1634, Ana_1635, Ana_1636, Ana_1637, Ana_1638, Ana_1639, Ana_1640, Ana_1641, Ana_1642 and Ana_1644. These genes showed the consistent homologies with each other except for a few micro-organisms. OT170 has full gene lengths of Ana_1634, Ana_1635, Ana_1636, Ana_1637, Ana_1642 and Ana_1644 but Ana_1638, Ana_1639, Ana_1640 and Ana_1641 genes were lacking. S24V (an *A. oris* isolate) has also a half gene length for Ana_1641 as compared to other *A. oris* isolates (134 versus 472). MMRCO6-1 was found to contain half gene length of Ana_1637, which is 333 amino acid (a.a) as compared to gene length (Ana_1637) of 737 a.a of other *A. oris* isolates. OT171 has 169 amino acid sequence length for Ana_1642 while all other *A. oris* isolates has 281 amino acid sequence length. Similarly M48-1B-1 has 598 sequence length for Ana_1637 while standard a.acid sequence length for this gene in all other *A. oris* isolates was 737 a.a. The percentage homology observed falls in the range of 84 to 100 % for all of the strains. *A. naeslundii* isolates showed very inconsistent amino acid sequence length and homologies with *A. oris*-MG1 for the gene regions of Ana_1634 to Ana_1644. Ana_1634, Ana_1635 and Ana_1644 genes were complete for all of the *A. naeslundii* isolates. W8-2-3 only has a gene length of 222 a.a for the Ana_1635 gene as compared to a gene length of 407 a.a in the other isolates. Ana_1636 showed a massive degradation and has an amino acid sequence length of 88 for R19039, R13240, F12B1, WE6B-3, NCTC 10301 and S43L; 119 for MB-1, T23P-1, R8152, while Pn6N and CCUG 37599 have the maximum length of 147 and finally MMRC12-1 had a 72 amino acid sequence length. There is no gene sequence found for Ana_1637, Ana_1638, Ana_1639, and Ana_1640 in any of the *A. naeslundii* isolates.

Similarly for gene Ana_1641, there was extensive gene degradation was observed and only five of the isolates contained a gene length of 99 with 81-83% homology (F6E1, Pn6N, MMRC12-1, and CCUG 37599). The rest of the genes, Ana_1642 and Ana_1644, contained complete gene sequences with homologies of 84-86%.

Table 5.2: Amino acid sequence length and their percentage homology to *A. oris*-MG1

Strains	ANA_1634 (1451)	ANA_1635 (1452)	ANA_1636 (1453)	ANA_1637 (1454)	ANA_1638 (1456)	ANA_1639 (1457)	ANA_1640 (1458)	ANA_1641 (1459)	ANA_1642 (1460)	ANA_1644 (1461)
A. oris										
*MG1	293	432	404	**737**	**431**	**318**	**288**	473	281	422
G53E	293, 91	407, 91	404, 94	**737, 97**	**431, 99**	**318, 99**	**288, 99**	473, 98	281, 96	422, 94
*OT175	293, 99	432, 99	404, 94	**737, 98**	**431, 98**	**318, 99**	**288, 99**	473, 98	281, 95	422, 94
W11-1-1	293, 97	410, 93	404, 93	**737, 97**	**431, 98**	**318, 98**	**288, 98**	473, 98	281, 95	422, 95
F4D1	293, 97	410, 93	404, 97	**737, 98**	**431, 99**	**318, 99**	**288, 98**	473, 98	282, 94	422, 95
R23275	293, 94	407, 90	404, 93	**737, 97**	**487, 98**	**318,100**	**288, 98**	473, 99	281, 94	422, 91
A7A-1	293, 90	407, 90	404, 95	**737, 97**	**431, 98**	**318, 99**	**288, 99**	473, 98	281, 97	422, 95
*C505	293, 90	411, 90	443, 94	**737, 98**	**431, 99**	**318, 99**	**288, 99**	473, 99	281, 98	414, 96
*K20	293, 99	429, 97	404, 97	**737, 98**	**431, 100**	**318, 99**	**288, 99**	473, 99	281, 97	**X
WE8B-23	293, 92	426, 95	404, 91	**737, 96**	**431, 98**	**318, 98**	**288,97**	473, 97	281, 94	422, 94
R11372	293, 99	426, 95	404, 91	**737, 97**	**431, 97**	**318, 98**	**288, 98**	473, 97	281, 95	422, 94
A19A-1	293, 99	426, 95	404, 92	**737, 97**	**431, 97**	**318, 98**	**288, 97**	473, 97	281, 94	422, 94
CCUG 34286	293, 99	426, 95	404, 91	**737, 97**	**431, 97**	**318, 98**	**288, 97**	473, 97	281, 96	422, 94
F28B1	293, 99	426, 95	404, 93	**737, 97**	**431, 98**	**318, 98**	**288, 98**	473, 97	281, 95	422, 95
M48-1B-1	293, 99	426, 95	404, 92	**598, 97**	**431, 98**	**318, 98**	**288, 98**	473, 97	281, 96	422, 94
MMRCO6-1	293, 98	426, 95	404, 92	**333, 98**	**431, 98**	**318, 99**	**288, 98**	473, 97	281, 96	422, 93
P6N	293, 91	443, 92	404, 88	**737, 95**	**431, 96**	**331, 93**	**288, 97**	473, 94	277, 92	422, 90
CCUG33920	293, 94	440, 92	435, 89	**737, 95**	**431, 97**	**331, 94**	**288, 98**	473, 96	297, 91	422, 89
*OT171	293, 94	440, 93	435, 89	**737, 94**	**431, 97**	**331, 94**	**288, 99**	473, 96	**169, 92**	423, 90
S24V	293, 91	426, 94	404, 91	**737, 96**	**431, 98**	**318, 99**	**288, 97**	**134, 96**	281, 98	403, 93
R21091	293, 94	426, 94	404, 90	**737, 96**	**431, 97**	**318, 98**	**288, 97**	473, 95	281, 94	422, 93
S64C	293, 94	426, 94	404, 90	**737, 96**	**431, 98**	**318, 99**	**288, 97**	473, 97	281, 95	422, 91
*OT170	293, 90	434, 91	340, 81	**708, 91**	**X	**X	**X	**X	277, 84	422, 84

A. naeslundii	Ana_1634	Ana_1635	Ana_1636	Ana_1637	Ana_1638	Ana_1639	Ana_1640	Ana_1641	Ana_1642	Ana_1644
CCUG 35334	293, 89	407, 87	-	-	-	-	-	-	273, 85	422, 84
G127B	293, 89	407, 87	-	-	-	-	-	-	273, 85	415, 85
MB-1	293, 89	407, 87	119, 78	-	-	-	-	-	276, 85	422, 84
NCTC 10301	293, 89	407, 87	88, 85	-	-	-	-	-	273, 86	422, 84
R24330	293, 89	407, 87	-	-	-	-	-	217, 61	280, 83	415, 84
S43L	293, 89	411, 87	88, 85	-	-	-	-	-	276, 83	415, 84
S44D	293, 88	407, 87	-	-	-	-	-	-	276, 85	415, 84
S65A	293, 89	407, 87	-	-	-	-	-	-	273, 86	415, 84
T23P-1	293, 89	407, 87	119, 78	-	-	-	-	-	276, 85	422, 84
CCUG 37599	293, 89	407, 87	147, 85	-	-	-	-	99, 83	280, 83	422, 83
F6E1	293, 89	407, 87	-	-	-	-	-	99, 81	280, 83	403, 85
Pn6N	293, 89	407, 87	147, 85	-	-	-	-	99, 83	280, 83	415, 84
MMRC12-1	293, 89	407, 87	72, 88	-	-	-	-	99, 81	280, 83	415, 84
F12B1	293, 89	407, 87	88, 84	-	-	-	-	-	280, 83	415, 84
R8152	293, 89	407, 87	119, 82	-	-	-	-	-	280, 83	415, 84
R13240	293, 89	407, 87	88, 85	-	-	-	-	-	276, 86	415, 84
R19039	293, 89	407, 87	88, 80	-	-	-	-	-	280, 83	415, 84
W8-2-3	293, 89	222, 84	-	-	-	-	-	-	273, 84	422, 84
WE6B-3	293, 89	407, 87	88, 85	-	-	-	-	-	276, 86	415, 84

***** Sequences in public databases, all are *A. oris*
* Ana_1451-Ana_1461 is the gene numbers in RAST after trimming of the denovo sequences
* LNB operon is marked red (Ana_1637 to Ana_1640)
* ANA_1634 until ANA_1644 are gene locus of *Actinomyces oris*-MG1 in CMR website and X = not known.
* Values shown are amino acid length and percentage homology to *A. oris*-MG1

5.4 Identification Of Lacto-*N*-Biose Gene Cluster Using PCR Based Demonstration

5.4.1 Material And Methods

Preliminary analysis suggested a gene cluster similar to the Lacto-*N*-Biose operon (Section 5.3) was present in *A oris* strains but absent in *A. naeslundii*. To investigate the presence of Lacto-*N*-Biose gene cluster in *A. oris* and *A. naeslundii*, 25 additional isolates of each species were selected for comparison (Table 5.3), in addition to the strains subjected to genome sequencing. In addition 6 other genomes of Actinomyces phylotypes, clearly related to *A. oris* and *A. naeslundii*, on the basis of 16S sequencing were also investigated. These were MG1, c505, k20, OT170, OT171 and OT175. The LNB operon in *A. oris*-MG1 is shown in Figure 5.1 for the reference.

5.4.1.1 DNA Extraction

Cultures of freeze-dried isolates of 50 *Actinomyces* (Table: 5.3) grown on FAA plates for 48 hours were obtained. These strains were checked for their purity by Gram staining and single cell streaking on FAA plates. The pure colonies were obtained. Genomic DNA was prepared for all isolates using Proteinase K (Aas *et al*., 2005). Bacterial cells were suspended in 70 µl 2M NaCl solution using 1.5 ml centrifuge tubes to remove the extracellular polysaccharides. The tubes were spun at 10,000 rpm for 6 min. The supernatant was removed and pellets were resuspended in 50 µl TE buffer containing 0.5 % Tween® 20 (Tris-EDTA buffer, pH 8.0). 200 µl of Proteinase K solution (10mg/ml stock solution) was added in the tube. The tubes were incubated in a microtherm at 55 ºC for 2 h and then heated at 95 ºC for 5 min in order to inactivate the Proteinase K. The tubes were then centrifuged at 13,000 rpm for 1 min, the supernatants were transferred into new 1.5ml micro-centrifuge tubes and the DNA extracts were stored at -20 ºC.

Table 5.3: List of strains used in PCR based studies to observe Lacto-*N*-Biose operon region

A. oris		***A. naeslundii***	
Strain numbers	Strain name	Strain number	Strain name
1	S62B	**26**	E1-20
2	F5A1	**27**	M42-1-1
3	M47-1-1	**28**	T17-3
4	F24C1	**29**	Pn16E
5	U149-1	**30**	R24330
6	F23A1	**31**	T20P-1
7	MMRCO2-1	**32**	S44D
8	A18A-3	**33**	S49H
9	F2A12	**34**	CCUG34725
10	F21A1	**35**	R9841
11	WE10B-1	**36**	R19039
12	G4BB	**37**	S38H
13	A7A-1	**38**	S41C
14	S33A	**39**	U136-1
15	UN4B6c	**40**	R8152
16	R23277	**41**	F6E1
17	M46-1-1	**42**	G126D
18	S55N	**43**	Pn1GA
19	WE9A10-1	**44**	T14P-1
20	P5K	**45**	G127B
21	R14746	**46**	G51C
22	F7j1	**47**	T9P-1
23	F2A-1	**48**	S53C
24	F11C1	**49**	T23P-1
25	G140C	**50**	WE7B1

5.4.1.2 Primer Design (Primer Design Between Genes Ana_1636 to Ana_1642)

The primers were designed using the consensus gene sequences of 36 *Actinomyces oris* isolates. The conserved gene regions were transferred into BioEdit program version 2.0. and aligned using ClustalW. The primers chosen were of size 18-20 nucleotide, G+C mol, secondary structure and annealing temperatures were confirmed by using software **http://www.basic.northwestern.edu/biotools/oligocalc.html).** The PCR conditions used initially for each primer pair were initial denaturation at 94 °C for 10 minute, then 30 cycles of denaturation at 94 °C for 30 sec, Primer annealing at 49 °C for 30 sec, extension at 72 °C for 90 sec, and final elongation for 72 °C for 5 minute. Primers were designed between gene regions Ana_1636—Ana_1637, Ana_1638—Ana_1639, Ana_1640—Ana_1641, Ana_1641—Ana_1642, Ana_1642—Ana_1644, Ana_1635 and Ana_1637 used as a positive control for *A. oris* and between Ana_1636—Ana_1642 for *A. naeslundii* isolates.

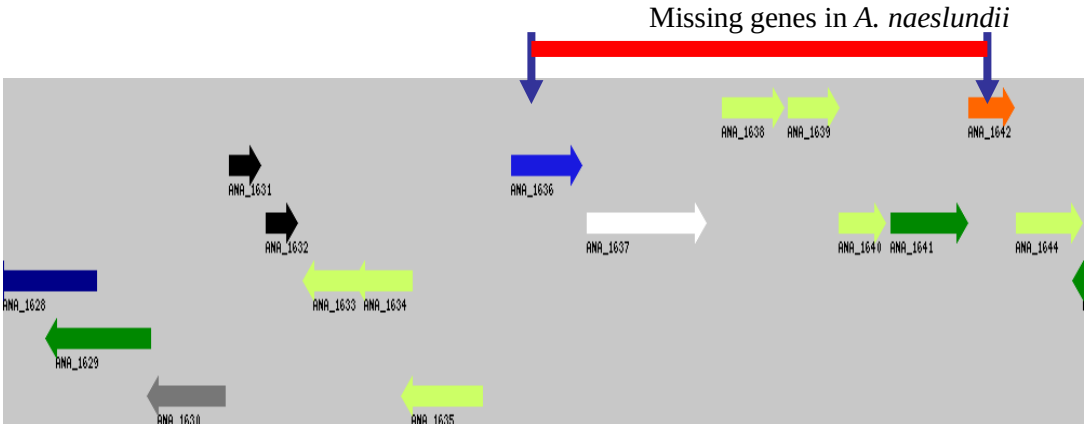


Figure 5.4: Schematic structure of the gene cluster between the ANA_1633 to ANA_1644

***(taken from the JCVI, CMR website) and the Primer design using *A. oris* - MG1: between arrows.*names of genes mentioned in Table 5.1**

5.4.1.3 Primers For *A. oris*

The primers used for sequencing the Lacto-*N*-Biose operon region in *Actinomyces oris* isolates were shown below. The PCR conditions were given for each primer combinations. The PCR products were visualized on a 1% agarose gel stained with Gel-red™ (10,000x in DMSO; Biotium, Hayward, CA, USA). The DNA fragments were visualized with a UV transilluminator and the gel images were recorded with an Alpha Fluorchem ™ FC8800 (Alpha Innotech Corporation).

5.4.1.3.1 Sequencing Of ANA_1635

Primer	Primer Sequence
1635 F2	ACAAGGGCGAGATCTT
1635 R2	ATCGTCTTGCCGAGCTT

The PCR programme used was (a) 94 ºC for 10 min, (b) 94 ºC for 30 sec, (c) 49 ºC for 30 sec (d)72 ºC for 90 sec, go to (b) 30 times (e)72 ºC for 5 min.

5.4.1.3.2 Sequencing Of ANA_1637

Primer	Primer Sequence
1637 F2	CAACTACCTCACCAAC
1637 R3	GAAGAAGTACGGCAGGAA

The PCR programme used was (a) 94 ºC for 10 min, (b) 94 ºC for 30 sec, (c) 49 ºC for 30 sec (d)72 ºC for 90 sec, go to (b) 30 times (e) 72 ºC for 5 min.

5.4.1.3.3 Sequencing Between ANA_1636 to ANA_1637

Primer	Primer Sequence
1636 F4	TCCGTGCTCGACAA
1636 F5	TGGATGAGCGCATCGA
1637 R5	AGCTCCGAGACGAT
1637 R7	AAGCTGACGGTGTA

The PCR programme was (a) 94 ºC for 10 min, (b) 94 ºC for 30 sec, (c) 49 ºC for 30 sec (d)72 ºC for 90 sec, go to (b) 30 to 40 times (e)72 ºC for 5 min.

5.4.1.3.4 Sequencing Between ANA_1638 To ANA_1639

Primer	Primer Sequence (5′→3′)
1638 F2	**CCATCTCCACCTTCTT**
1639 R1	**GAACATGGCCAGGAA**

The PCR programme used was (a) 94 ºC for 10 min, (b) 94 ºC for 30 sec, (c) 49 ºC for 30 sec (d)72 ºC for 90 sec, go to (b) 34 times (e)72 ºC for 5 min.

5.4.1.3.5 Sequencing Between ANA_1640 To ANA_1641

Primer	Primer Sequence (5′→3′)
1640 F1	GTCGACCTACTTCAAC
1641 R2	GATGCCGTTGTACAG

The PCR programme used was (a) 94 ºC for 10 min, (b) 94 ºC for 30 sec, (c) 49 ºC for 30 sec (d)72 ºC for 90 sec, go to (b) 34 times (e) 72 ºC for 5 min.

5.4.1.3.6 Sequencing Between ANA_1641 To ANA_1642

Primer	Primer Sequence (5′→3′)
1641 F5	GGCTGCAAGGTCTA
1642 R7	GCTGTAGTCGATGA

The PCR programme used was (a) 94 ºC for 10 min, (b) 94 ºC for 30 sec, (c) 47 ºC for 30 sec (d)72 ºC for 90 sec, go to (b) 35 times (e)72 ºC for 5min.

5.4.1.3.7 Sequencing Between ANA_1642 To ANA_1644

Primer	Primer Sequence (5′→3′)
1642 F2	CTGGAGGAGATCGT
1644 R2	CCGAAGTTGAGGGT

The PCR programme used was (a) 94 ºC for 10 min, (b) 94 ºC for 30 sec, (c) 49 ºC for 30 sec (d)72 ºC for 90 sec, go to (b) 34 times (e)72 ºC for 5 min**.**

5.4.1.4 Primers For *A. naeslundii*

The primers were choosen for *A. naeslundii* isolates to confirm the presence of the Lacto-*N*-Biose operon region by desigining primers between genes Ana_1636 to Ana_1642.

5.4.1.4.1 Sequencing Between ANA_1636 to ANA_1642

Primer	Primer Sequence (5′→3′)
1636 F1	TGCGCATCTCGAACAT
1636 F9	GTCACTCATTCTGCG
1642 R2	AAGTCCTCGTGCTC
1642 R7	GCTGTAGTCGATGA

The following primer combinations were used with PCR programs**.**

a) Ana_1636-F1 to Ana_1642-R2

The PCR programme used was (a) 94 ºC for 10 min, (b) 94 ºC for 30 sec, (c) 49 to 51 ºC for 30 sec (d)72 ºC for 90 sec, go to (b) 34 to 40 times (e)72 ºC for 5 min

b) Ana_1636 F9 to Ana_1642 R2

The PCR programme used was (a) 94 ºC for 10 min, (b) 94 ºC for 30 sec, (c) 49 ºC for 30sec (d)72 ºC for 90sec, go to (b) 34 times (e)72 ºC for 5min

5.4.2 Results

To confirm the presence of Lacto-*N*-Biose gene cluster in other isolates of *Actinomyces*, 50 more isolates were selected from *A. oris* and *A. naeslundii* species (Table 5.4). The sets of gene numbers starting from ANA_1635 to ANA_1644 were chosen based on their presence (conserved) in *A. oris* and absence in *A. naeslundii* strains during preliminary studies of gene by gene analysis (Figure 5.4). Initially the problem was encountered while designing the primers due to non-availability of much genomic data on the internet database for *A. oris* and *A. naeslundii* isolates but in early gene-by-gene investigation studies, sequences were obtained to design the primers from sequenced genomes using Illumina and Roche 454 in the current study. A region of the LNB gene was amplified by PCR (Figure 5.5) using the genomic DNA of *Actinomyces* as a template and a pair of primers (Table 5.4) that were designed based on the amino acid sequence of the gene operon responsible for utilization of LNB from genomic sequences of *Actinomyces* strains. Table 5.4 shows different primer combinations that were positive for *A. oris* strains (MG1, S62B) and *A. naeslundii* (NCTC 10301) strain. Few of the primer sets were selected for the sequencing of isolates. All DNA fragments were sequenced with a forward and reverse primer. The DNA sequences were analysed, trimmed and aligned using BioEdit software version 2.0.0 (Hall *et al.*, 1999). The beginning and ends of sequences were cut, and subsequently, the contig assembly programme (Gevers *et al.*, 2012) was used in order to align forward and reverse sequences of isolates.

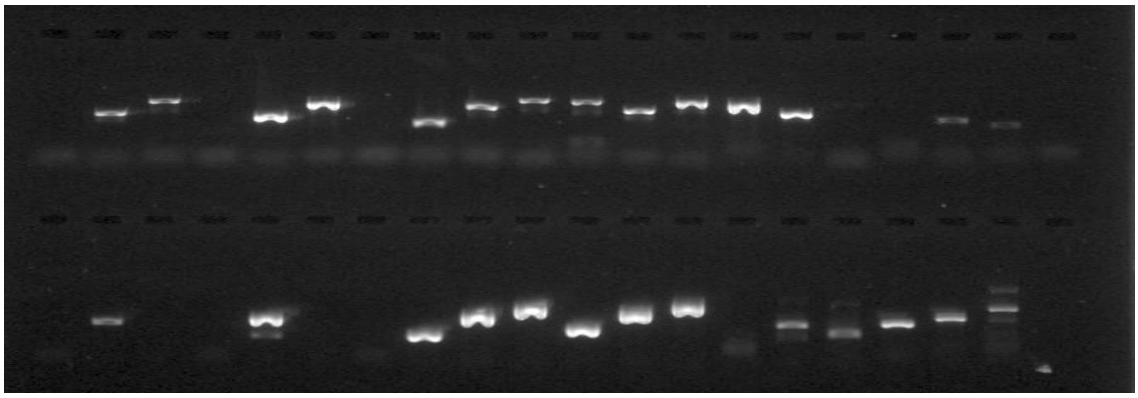


Figure 5.5: PCR of *A. oris*-MG1 isolate with primer combination from A-O1 (Table 5.4).

Table 5.4: Primer design in *A. oris*-MG1 and *A. naeslundii*-NCTC 10301

Primer for 1635			Primer for 1637			Primer 1638-1639		
No.	**Primers**	**Products**	**No.**	**Primers**	**Products**	**No.**	**Primers**	**Product**
A	1635 F1 -1635 R1	_	J	1637 F1 -1637 R1	+	S	1638 F1 -1639 R1	+
B	1635 F1 -1635 R2	+	K	1637 F1 -1637 R2	+	T	1638 F1 -1639 R2	_
C	1635 F1-1635 R3	+	L	1637 F1-1637 R3	+	U	1638 F1-1639 R3	-
D	1635 F2-1635 R1	-	M	1637 F2-1637 R1	+	V	1638 F2-1639 R1	+
E	1635 F2-1635 R2	+	N	1637 F2-1637 R2	+	W	1638 F2-1639 R2	-
F	1635 F2-1635 R3	+	O	1637 F2-1637 R3	+	X	1638 F2-1639 R3	-
G	1635 F3-1641 R2	-	P	1637 F3-1641 R1	-	Y	1638 F3-1639 R1	+
H	1635 F3-1641 R3	+	Q	1637 F3-1641 R2	-	Z	1638 F3-1639 R2	-
I	1635 F3-1641 R1	+	R	1637 F3-1641 R3	+	A1	1638 F3-1639 R3	-
Primer for 1640-1641			**Primer for 1642-1644**			**Primers for 1636-1642**		
No.	**Primers**	**Products**	**No.**	**Primers**	**Products**	**No**	**Primers**	**Products**
B1	1640 F1-1641 R1	+	K1	1642 F1-1644 R1	-	T1	1636F1-1642R2	+
C1	1640 F1-1641 R2	+	L1	1642 F1-1644 R2	+			
D1	1640 F1-1641 R3	+	M1	1642 F1-1644 R3	+	**Primers for 1636-1637**		
E1	1640 F2-1641 R1	+	N1	1642 F2-1644 R1	+	U1	1636F5-1637R5	+
F1	1640 F2-1641 R2	+	O1	1642 F2-1644 R2	+	V1	1636F4-1637R7	+
G1	1640 F2-1641 R3	+	P1	1642 F2-1644 R3	+	W1	1636F4-1637R7	+
H1	1640 F3-1641 R1	-	Q1	1642 F3-1644 R1	_			
I1	1640 F3-1641 R2	-	R1	1642 F3-1644 R2	-			
J1	1640 F3-1641 R3	-	S1	1642 F3-1644 R3	-			

* Sequences of primers are listed in section 5.4.1.3 to 5.4.1.4

5.4.2.1 *A. naeslundii* Aligned Between Genes Ana_1636-Ana_1642

The sequenced data was extracted using the RAST programme where the full genomes were uploaded and nucleotide sequences were extracted for Ana_1636 DNA with the flanking 1000 bases. Then FASTA files were copied into Bioedit files and aligned using ClustalW programme. There is an extensive degradation observed between gene regions ANA_1636 [xylose repressor] to ANA_1642 [glucose-1-phosphate thymidyltransferase] in 19 *A. naeslundii* isolates (Table 2.1).

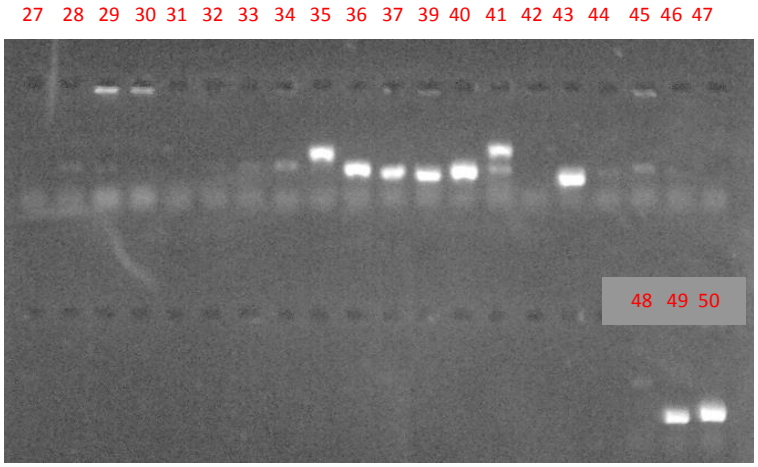


Figure 5.6: *A. naeslundii* isolates showed positive bands with primer combination 1636F1-1642R2.

*** Strains names for 27-50 are mentioned in Table 5.3**

The 25 (isolate number 26-50) selected isolates (Table 5.3) were obtained from freezer stock cultures. *A. naeslundii* NCTC 10301 was used as a control and amplified region selected was Ana_1636-Ana_1642. The PCR programme used was mentioned earlier in this section (Section 5.4.1.4.1) and each was run on a gel. The primer pairs which gave positive bands were selected to amplify the regions in other *A. naeslundii* isolates. A lot of effort was made to get the DNA bands on electrophoresis using different primer combinations. Not all 25 isolates gave positive bands with single primer combinations

due to extensive gene loss seen in the LNB operon region in 19 sequenced isolates therefore it was not very easy to amplify the region. R9841, R19039, S38H, U136-1, R8152, Pn1GA, T23P-1, WE7B1 (8 isolates) appeared positive (Figure 5.6) using Ana_1636-F1 to Ana_1642-R2 with PCR program while using the same primer combination stringency condition in PCR program was raised by 2 ºC to get more positive bands for the rest of the isolates which appeared as weak bands as shown in Figure 5.7. The strains which appeared positive were T17-3, R24330, S49H, CCUG 34725, T14P-1, G127B, and G51C (7 isolates) while S44D and S53C were positive when run using the PCR program with 40 cycles. The remaining 6 strains (E1-20, T20P-1, S49H, S41C, F6E1, and G126D) appeared positive with Ana_1636-F9 to Ana_1642-R2 primer combination using the PCR programe while T9P-1, S53C did not appear positive with any of the primer combinations tested.

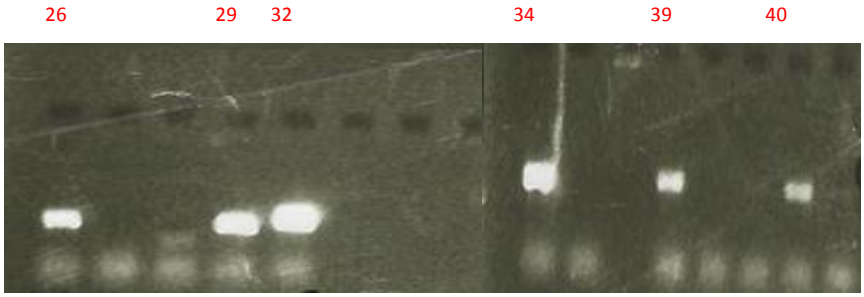


Figure 5.7: DNA bands appeared positive with the strains numbers (26-40)

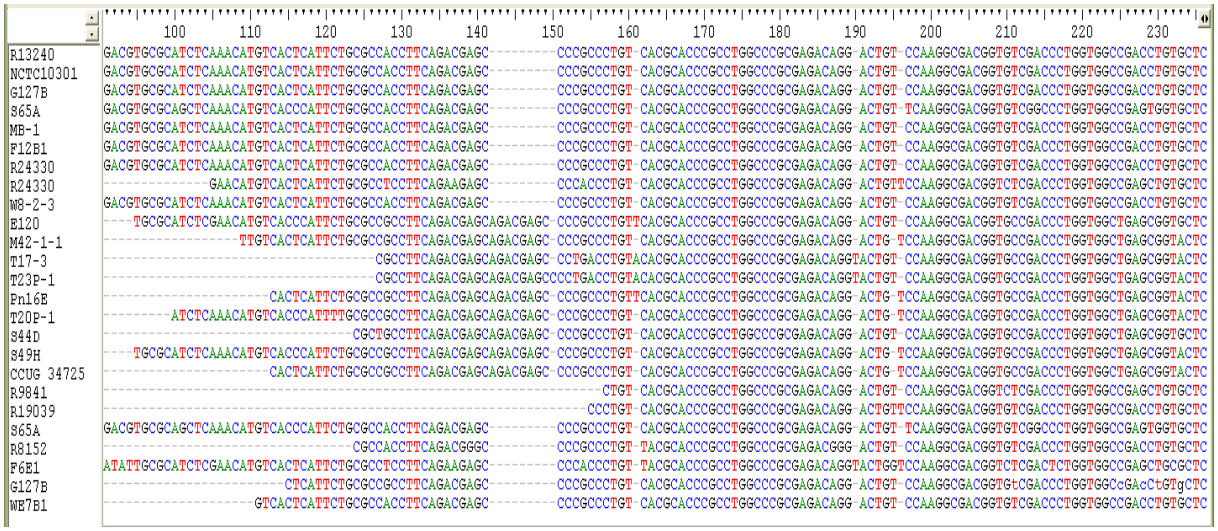
*Strain names are mentioned in Table 5.3

Similarly genomes (c505, K20, MG1, OT170, OT171, OT175) from the internet database were uploaded using the RAST programme and interesting sequences were extracted and added to the aligned sequences of 50 isolates obtained using PCR-based sequencing of the LNB operon. There were three types of alignment observed while aligning all of the sequences using BioEdit version 2.0.

5.4.2.1.1 Type I Sequences: 32 Strains

The Type I was observed in 32 *A. naeslundii* isolates when aligned between Ana_1636-Ana_1642 gene regions. Type 1 is shown in Figure 5.8. The alignment contained a gene region homologus to the 3' end of the Xylose Repressor (Ana_1636) and has 1-110 residues. This is followed by a region of residual sequences which did not come up in Blast results for the intergenic region. The other end contained a gene region homologous to Glucose-1-phosphate thymidyltransferase (Ana_1642) at 5' end and it contained 700-1100 nucleotides. The isolates which aligned properly with Type I are 15 *A. naeslundii* isolates from the sequenced database named G127B, MB-1, R24330, T23P-1, S44D, R19039, R8152, W8-2-3, S65A, R13240, CCUG 35334, F12B1, WE6B-3, NCTC 10301 and S43L and 13 from PCR sequences E1-20, M42-1-1, T17-3, Pn16E, T20P-1, S49H, CCUG 34725, R9841, R19039, R8152, F6E1, G127B, WE7B1. c505, MG1, k20 and OT171 were included from internet database. Only 25 aligned sequences are listed in Figure 5.8.

Type I: 32 strains



Xylose Repressor (Ana_1636) 3' end only

Intermediate region

Glucose-1-phosphate thymidyltransferase Ana_1642) 5' end only

Figure 5.8: *A. naeslundii* strains aligned between genes Ana_1636-Ana_1642

*** Screen shot of 25 aligned sequences**

5.4.2.1.2 Type II Sequences: 10 strains

Type II strains have the gene Xylose repressor (Ana_1636) at the 3' end and the Glucose-1-phosphate thymidyltransferase (Ana_1642) gene region at the 5' end, which is complete. In addition to the intermediate region there is a GlcNAc Phosphomutase (Ana_1641) present at the 5' end, while the intermediate region gave nothing by Blast comparisons (Figure 5.9). The Type II strains were F6E1, Pn6N, MMRC12-1, CCUG 37599, S38H, U136-1, Pn1GA, T14P-1, G51C, and T23P-1.

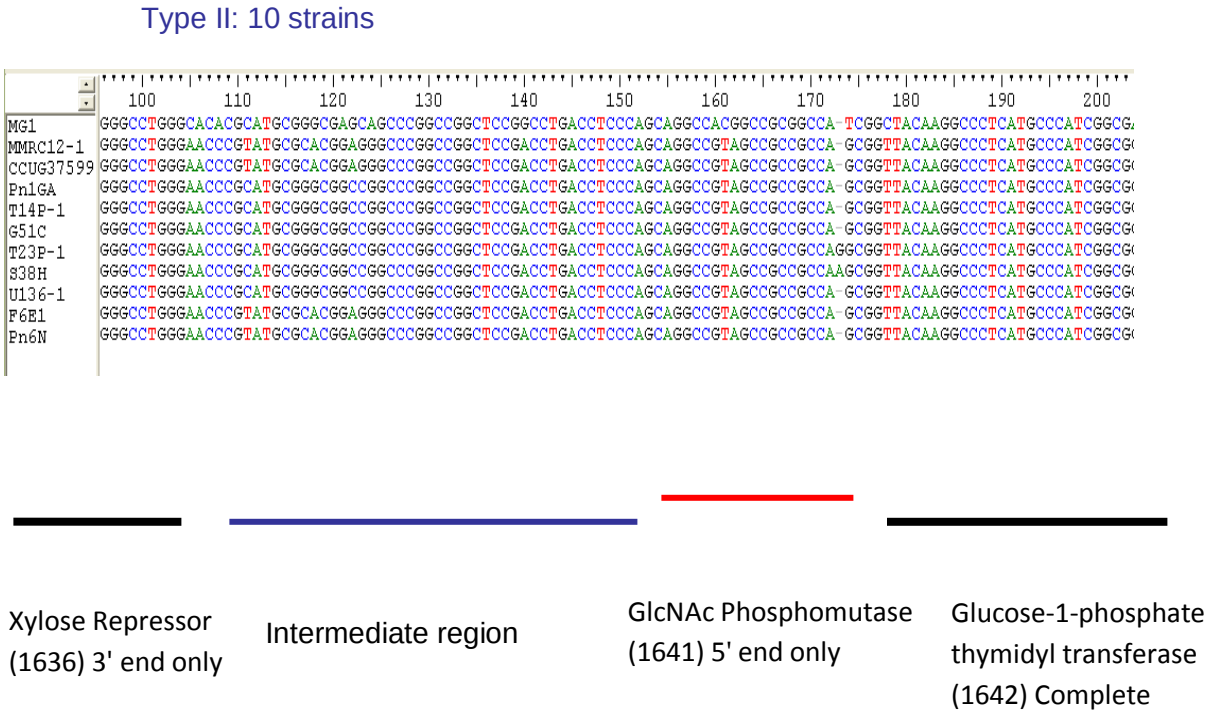


Figure 5.9: Screenshot of *A. naeslundii* Type II alignment between genes Ana_1636-Ana_1642

5.4.2.1.3 Type III Sequences (1 strain)

The third type of alignment designated as Type III was observed in only one strain. The Type III strain contains the LNB phosphorylase region (Ana_1637) at the 3' end in addition to the Xylose repressor (Ana_1636), GlcNAc Phosphomutase (Ana_1641) and the Glucose-1-thymidyltransferase (Ana_1642) (Figure 5.10).

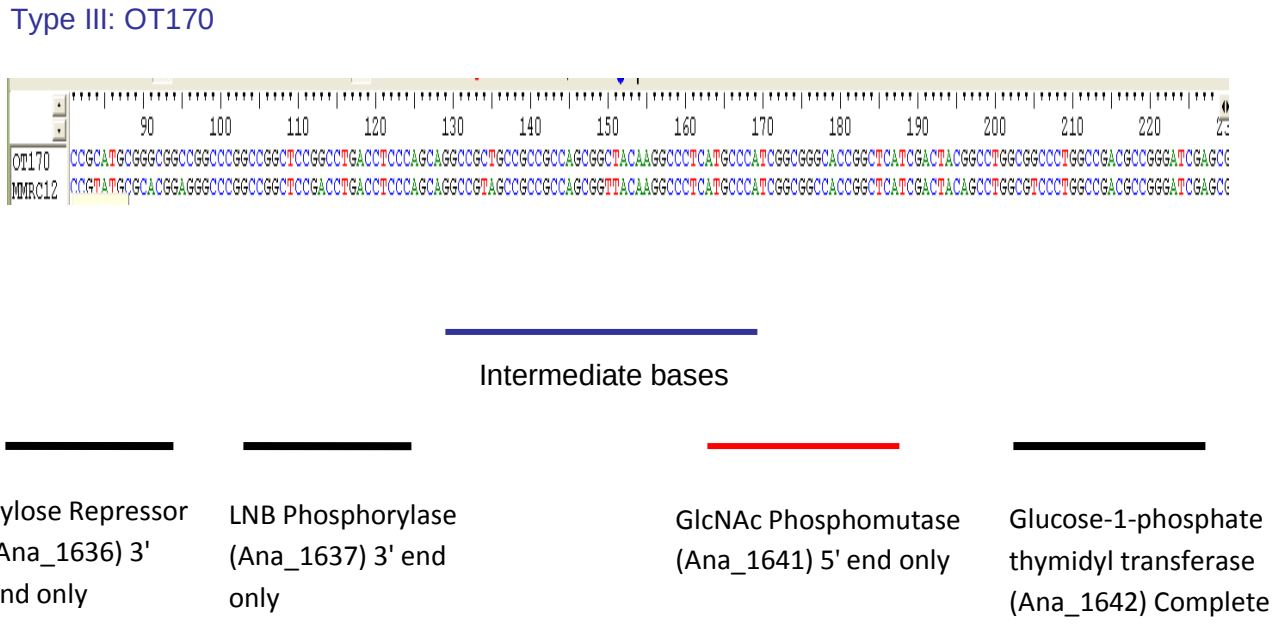


Figure 5.10: *A. naeslundii* aligned between Ana_1636-Ana_1642 with OT170 (Type III).

5.4.2.2 *A. oris* Alignment Comparison

5.4.2.2.1 Alignment Between Genes Ana_1636 to Ana_1637

Similarly the DNA sequences of individual genes of *A. oris* isolates were aligned using Clustal Multiple Alignment with the FAST algorithm and saved as a fasta file. The DNA sequences of ANA_1636 (Xylose repressor:XylR) and ANA_1637 (Lacto-*N*-biose phosphorylase (EC 2.4.1.211) [1,3-beta-galactosyl-N-acetylhexosamine phosphorylase] of MG1, OT171, c505, OT175, OT170 with the sequences obtained with Illumina and Roche 454 of 17 *A. oris* strains listed in Table 2.1 were included. Sequences were extracted from the RAST and were put together with 25 DNA sequences of selected *A. oris* isolates obtained from PCR sequencing listed in Table 5.4. All 48 strains appeared homologus to each other and there is no gene loss observed for any of the isolates as shown in Figure 5.11. The screen shot of the gene alignment is given between two genes, Ana_1636 to Ana_1637, and the intermediate region. The intergenic region has the same sequences in all *A. oris* strains and these sequences did not correspond to any known sequences when subjected to BLAST searches. Therefore, the presence of complete Ana_1636 and Ana_1637 genes was observed.

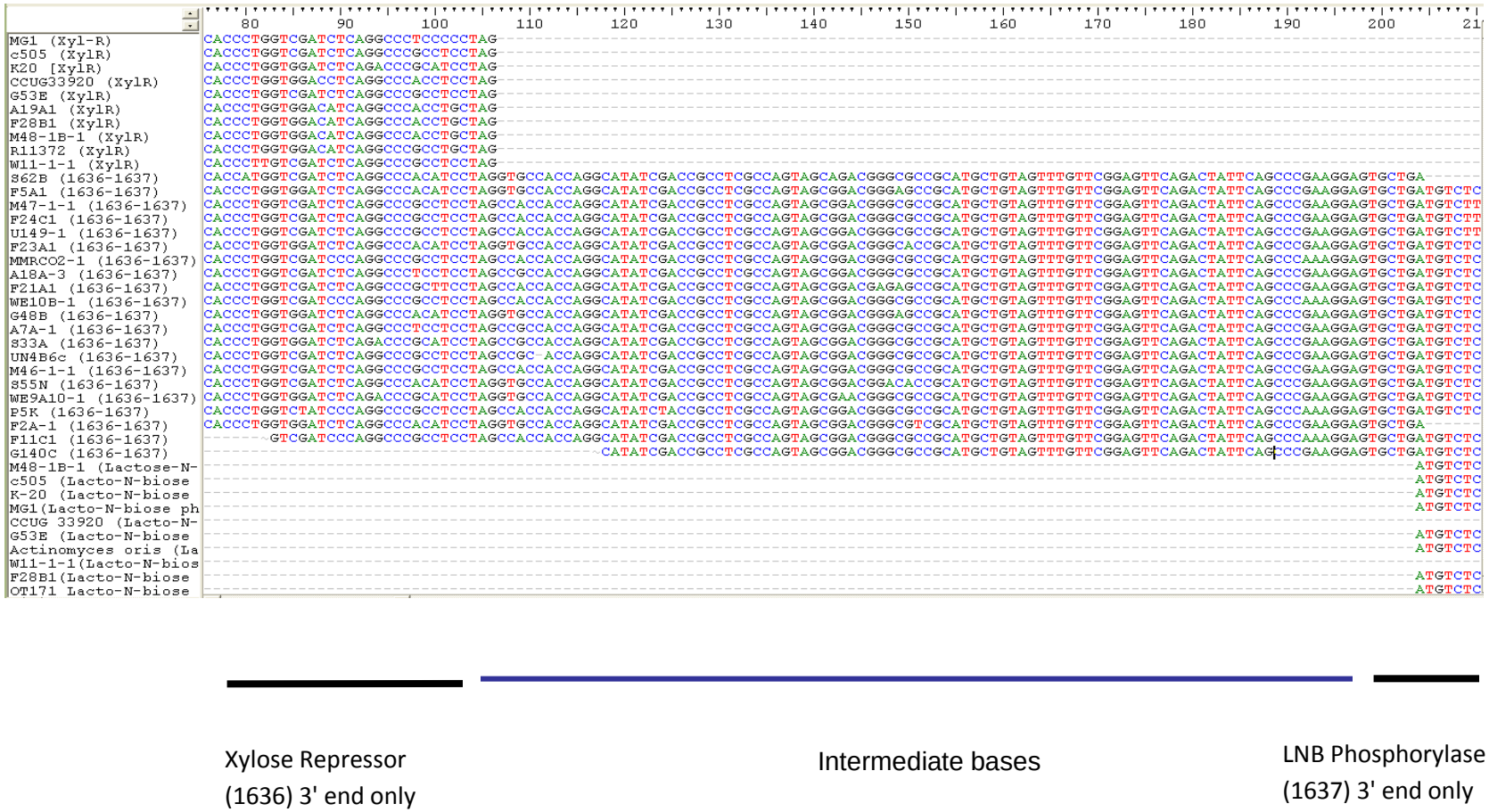


Figure 5.11: Screen shot of PCR sequences showing the intergenic region between ANA_1636 and ANA_1637 genes in the *A. oris* isolates sequenced in this study.

5.4.2.2.1.a Ana_1636 (Xylose Repressor) Aligned With c505, CCUG 33920 And OT171

There were few *A. oris* isolates which showed additional nucleotides at the 3' end of the Ana_1636 gene. The first 93 nucleotides were additional in *A. oris*-c505 as compared to *A. oris*-MG1 when aligned using Bioedit. The homologous sequences aligned from nucleotides number 93 to 1504. Similarly CCUG 33920 and OT171 contain additional nucleotides at the 3' end as shown in Figure 5.12, 5.13, 5.14.

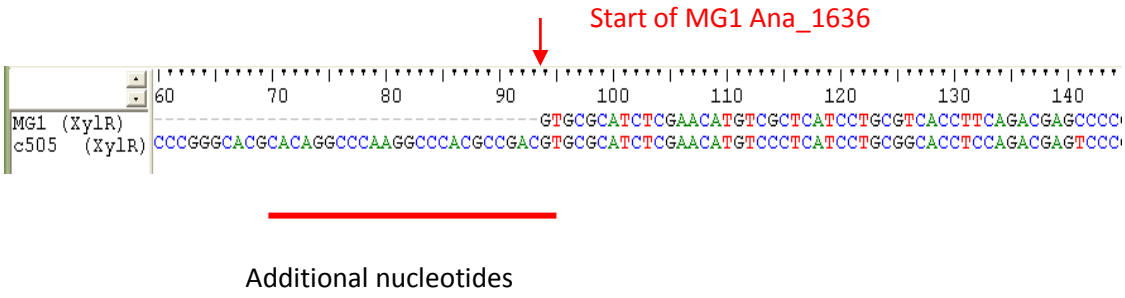


Figure 5.12: *A. oris*-MG1 gene Ana_1636 (XylR) aligned with c505

Figure 5.13: *A. oris*-MG1 gene Ana_1636 (XylR) aligned with CCUG 33920

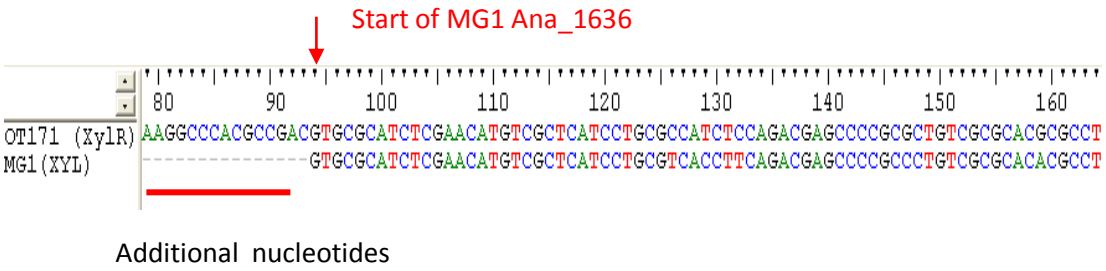
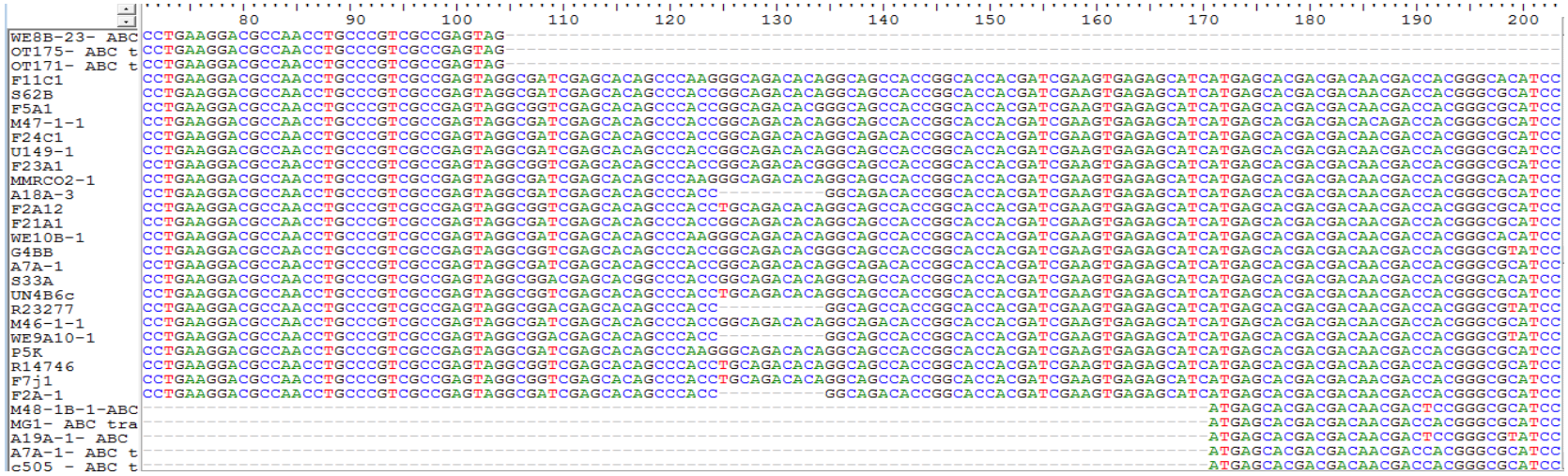


Figure 5.14: *A. oris*-MG1 gene Ana_1636 (XylR) aligned with OT171

5.4.2.2.2 Alignment Between Genes Ana_1638 to Ana_1639

A. oris isolates were aligned between gene numbers Ana_1638 to Ana_1639. The BioEdit files were prepared using sequences extracted from RAST of all of the Illumina and Roche 454 combined denovo sequences. The sequences were aligned using Clustal Multiple alignment tool with the FAST algorithm and saved as a fasta file. The DNA sequences of Ana_1638 (Predicted galacto-*N*-biose-/lacto-*N*-biose, ABC transporter, periplasmic substrate-binding protein) and Ana_1639 (Predicted galacto-*N*-biose-/lacto-*N*-biose, ABC transporter, permease component I) of MG1, OT171, c505, OT175, OT170 with the sequences obtained with Illumina combined with Roche 454 of 17 *A. oris* strains listed in Table 2.1. Sequences were extracted from RAST and were put together with the DNA sequences obtained from PCR sequencing of the 25 *A. oris* isolates listed in Table 5.3. All 47 strains appeared homologus to each other and there was no gene loss observed for any of the isolates, as shown in Figure 5.15. The screen shot of the gene alignment is given between two genes Ana_1638 to Ana_1639 and the intermediate region in Figure 5.15. The intergenic region has the same sequences in all *A. oris* strains and these sequences did not correspond to any known sequences when subjected to BLAST searches. Therefore, the presence of the complete Ana_1638 and Ana_1639 genes was observed.



Ana_1638 (Predicted galacto-*N*-biose-/lacto-*N*-biose, ABC transporter, periplasmic substrate-binding protein)

Intermediate Nucleotides

Ana_1639 (Predicted galacto-*N*-biose-/lacto-*N*-biose, ABC transporter, permease component I)

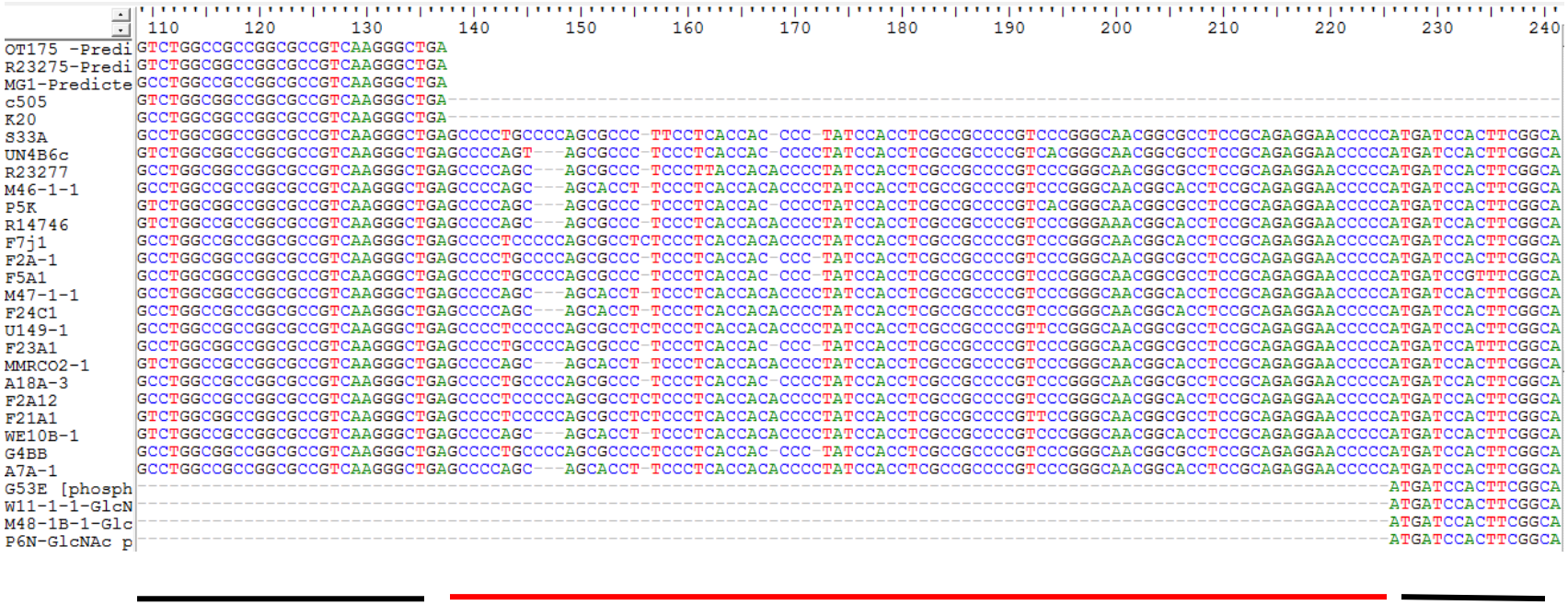
Figure 5.15: Screen shot of PCR sequences showing the intergenic region between ANA_1638 and ANA_1639 genes in the *A. oris* isolates sequenced in this study. *31 of 47 strains were shown

5.4.2.2.3 Alignment Between Genes Ana_1640 to Ana_1641

Similarly *A. oris* isolates were aligned between gene numbers Ana_1640 (Predicted galacto-*N*-biose-/lacto-*N*-biose, ABC transporter, permease component 2) to Ana_1641 (GlcNAc Phosphomutase (EC 5.4.2.3)). All *A. oris* isolates appeared homologus to each other and there was no gene loss observed for any of the isolates as shown in Figure 5.16. The screen shot of the gene alignment is given between genes Ana_1638 to Ana_1639 and the intermediate region in Figure 5.16. The intergenic region has the same sequences in all *A. oris* strains and these sequences did not correspond to any known sequences when subjected to Blast searches. Therefore the presence of the complete Ana_1640 and Ana_1641 genes was observed

5.4.2.2.4 Alignment Between Genes Ana_1642 to Ana_1644

A. oris isolates were aligned between gene numbers Ana_1642 (glucose-1-phosphate thymidylyltransferase (EC: 2.7.7.24)) to Ana_1644 (Galactokinase (EC 2.7.1.6)). All *A. oris* isolates appeared homologus to each other and there was no gene loss observed for any of the isolates as shown in Figure 5.17. The screen shot of the gene alignment was given between two genes Ana_1642 to Ana_1644 and intermediate region in Figure 5.17. The intergenic region has the same sequences in all *A. oris* strains and these sequences did not correspond to any known sequences when subjected to BLAST searches. Therefore the presence of the complete Ana_1642 and Ana_1644 genes was observed.



Ana_1640 (Predicted galacto-*N*-biose-/lacto-*N*-biose, ABC transporter, permease component 2)

Intermediate Bases

Ana_1641 (GlcNAc Phosphomutase (EC 5.4.2.3)

Figure 5.16: Screen shot of PCR sequences showing the intergenic region between ANA_1640 and ANA_1641 genes in the *A. oris* isolates sequenced in this study.

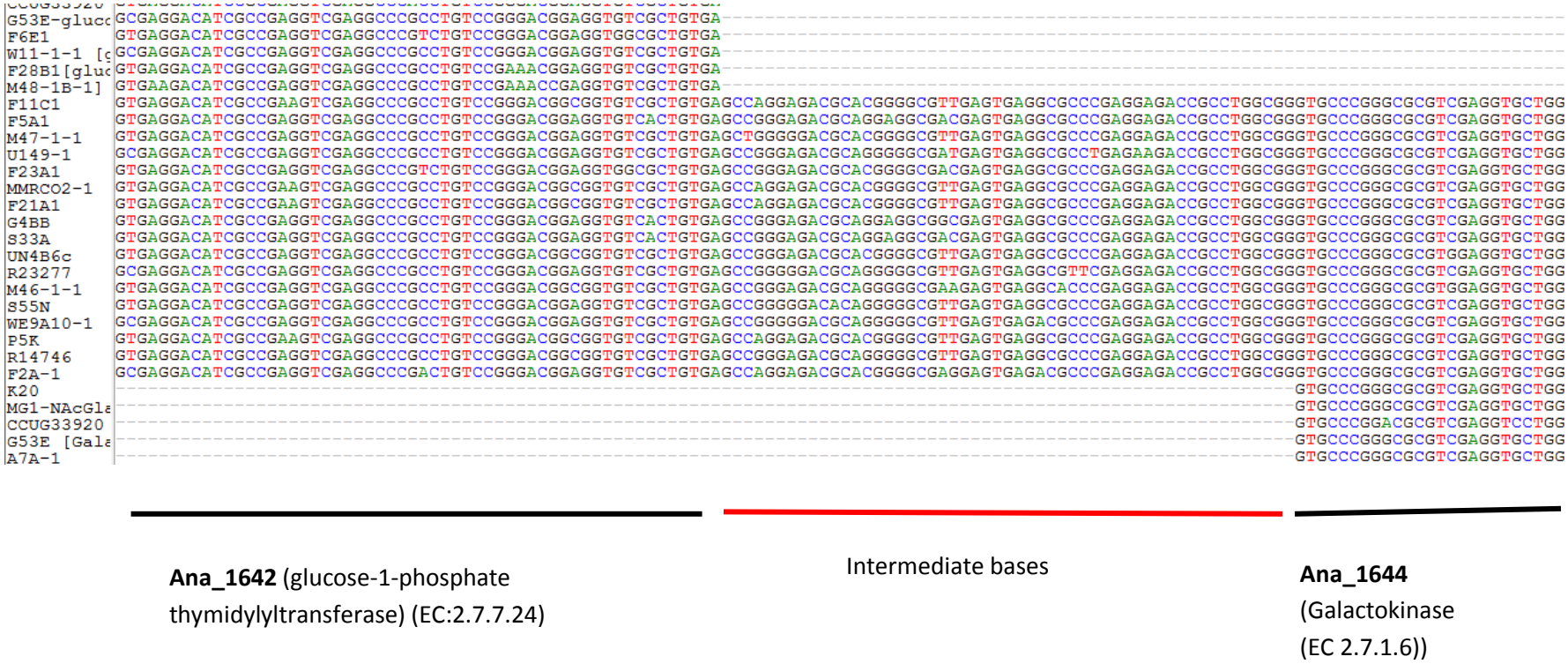


Figure 5.17: Screen shot of PCR sequences showing the intergenic region between ANA_1642 and ANA_1644 genes in the *A. oris* isolates sequenced in this study.

5.5 Reverse Transcriptase-PCR (RT-PCR) Demonstrates Up-Regulation Of The LNB Phosphorylase Gene (Ana_1637) When *A. oris* Was Grown In LNB

5.5.1 Material And Methods

5.5.1.1 Cultures

The cultures of *A. oris*- MG1 and *A. naeslundii*-NCTC 10301 were grown anaerobically at 37 ^{o}C on FAA plates supplemented with 5% (v/v) horse blood. The cultures were grown in Peptone Yeast Extract (PYE) (1:4) medium containing 1% glucose and 1% Lacto-*N*-Biose I separately and was incubated at 37 ^{o}C anaerobically for 24 hrs.

5.5.1.2 Primer Design

The gene sequence of Lacto-*N*-Biose Phosphorylase (Ana_1637) of *A. oris*-MG1 was downloaded from the 'Comprehensive Microbial Resource' (CMR) database of the J. Craig Venture Institute (JCVI; http;//cmr.jcvi.org/cgi-bin/CMR?GenomePage.cgi?org=gan). The Primers were designed using an online tool provided on the website http://www.basic.northwestern.edu/biotools/oligocalc.html. An optimal primer has a size of 18-22bp, a G+C mol% content of about 50-53%, an annealing temperature between 50 to 60 ^{o}C, no palindromes, no hairpin and no self dimer formation was checked through an online Oligo calculator. A pair of primers, forward and reverse, was selected to obtain a sequencing amplicon of 1017 bp of the Lacto-*N*-Biose Phosphorylase gene. The chosen primer sequences within *A. oris*-MG1 genome were tested for their specificity by BLASTing through the CMR website. The sequences used as forward and reverse primers are shown in Table 5.5.

Table 5.5: Primers for RT-PCR

Primer	Primer Sequence (5′→3′)
LNBP- F2	ACAACTACCTCACCAACG
LNBP-R2	AGGTAGGTGTAGGTCT

The dried primers were obtained from IDT (Integrated DNA Technologies, Invitrogen) and were rehydrated to a concentration of 100µM stock solution. 10µM working dilution was prepared from the stock solution and stored at -20 °C.

The PCR amplification of DNA extracts was carried out using the following program: 94 ^{o}C for 10 min (initial denaturation); 45 cycles of 94 ^{o}C for 30 sec (denaturation), 49 ^{o}C for 30 sec (Primer annealing) and 72 ^{o}C for 90 sec (elongation) and a last step at 72 ^{o}C for 5 min (final elongation). The PCR products were visualized on a 1% agarose gel stained with Gel-red^{TM} (10,000x in DMSO; Biotium, Hayward, CA, USA). The DNA fragments were visualized with a UV transilluminator and the gel images were recorded with an Alpha Fluorchem ^{TM} FC8800 (Alpha Innotech Corporation).

The cultures of *A. oris*-MG1 and *A. naeslundii*-NCTC 10301 were grown anaerobically at 37 ^{o}C overnight in 1.8 ml of PYE (1:4) medium containing 1% glucose in one tube and 1% Lacto-*N*-Biose in another tube. The 3.6 ml of RNAprotect®Bacteria Reagent (Cat NO: 76506) was added to the 1.8ml of cultures after 24 hours of growth. 20 mg of Lysozyme (L6876-5G, Sigma) and 5 mg of Achromopeptidase (A3547-1MU, Sigma) mixture was freshly prepared using RNase free water and was added to cultures to break up the cell wall of *Actinomyces oris* (MG1) and *A. naeslundii* (NCTC 10301). The mixture was incubated at 37 ^{o}C for 10 min.

5.5.1.3 RNA Extraction

Before the start of RNA extraction protocol, special arrangements were required due to the sensitivity of RNA to rapid degradation and contamination. The workbench was cleaned with Trigene II disinfectant (Cat No: CLE1312, SLS) and RNaseZap (Cat No: AM9780; AM9782, Ambion) as well as boxes, pipetts etc. The **MR2** vial was kept on ice during work. DNaseI enzyme (Qiagen) was taken out of the freezer and kept on ice. RDD buffer was taken out from fridge and left on bench to bring it to room temperature. The gloves were sprayed throughout the extraction protocol with RNaseZap before opening new tubes.

UltraClean^{TM} Microbial RNA isolation Kit (Cat No: 15800-50, MOBIO, Carlsbad, CA) was used to extract the RNA from cultures grown in glucose and oligosaccharides. 1.8ml of RNAprotect® and Lysis enzyme-treated microbial culture was added in a 2ml collection tube and centrifuged at 10,000 x *g* for 30 seconds. Supernatant was decanted and was completely removed with a pipette tip. The cell pellet was resuspended in 300 µl of Solution MR1 and vortexed gently to mix. The resuspended cells were transferred to a MicroRNA Bead tube. Then 15 µl of Solution MR2 was added to the MicroRNA Bead Tube and briefly vortexed. The tubes were heated at 65 ^{o}C for 10mins. The tubes were secured horizontally with tape on a MO BIO Vortex Adapter holder (MO BIO Catalogue No: 13000-V1) on a flat bed vortex pad. The tubes were centrifuged for 30sec at 10,000 *g*. The supernatant was transferred to a clean 2ml collection tube. 500µl of Solution MR3 (Lysis buffer) was added to the supernatant and vortexed for 5 sec. Then 250µl of Solution MR4 (Lysis buffer) was added and incubated at 4 ˚C for 5 min. The tubes were centrifuged for 1 min at 10,000 x *g*. The supernatant was transferred carefully to a clean 2ml collection tube. 650µl of the preparation was loaded onto a spin filter and centrifuged at 10,000 x *g* for 30 sec. The flow through was discarded. 300µl of Solution MR5 was added and centrifuged for 30 sec at 10,000 x *g*. The flow through was discarded.

5.5.1.4 RNAase Free DNase Treatment

The RNA can contain traces of genomic DNA therefore the samples were digested with RNAase free DNAase. The DNase Treatment was carried out on a spin column using RNase-Free DNase Set Kit (Cat No: 79254, Qiagen). DNase I incubation mix was prepared by adding 10 µl of DNase I in 70 µl of buffer RDD. The total of 80 µl of incubation mix was prepared per sample and poured onto the top of the spin column membrane. The mixture was left for 7 min at room temperature. The contents were centrifuged after 5 mins on pulse for 1 sec to bring the DNase I mixture down to the column. The tubes were spun for 10 sec to bring down the liquid after 2 mins.

Washing steps were carried out after DNase I treatment. 250 µl of 100% Ethyl alcohol (cold) was added to the column. The tubes were centrifuged at 8000g for 15 sec. The flow through was discarded. Then 500 µl of RPE buffer was added to the column. The tubes were centrifuged for 15 sec at 8000g. The flow through was discarded. 500 µl of 80% Ethyl alcohol was added to the column. The tubes were centrifuged for 20 sec at 8000g. The flow through was discarded. The spin column was placed in a new tube and spun for 2 min at 8000g. Then 50 µl x 2 of RNase free water was added on top of the filter for the elution of RNA. Elution was carried out at 8000g for 30sec and then for the second load at full speed for 30 sec. The tubes were flicked gently to mix and left on ice for 3 min and freeze at -80 ˚C for storage.

Normal PCR was carried out using housekeeping gene primers for MetG to check the purity of the RNA (for DNA contamination). The DNA sample was also included in the PCR run as a positive control and House keeping gene primers of MetG were used. The following PCR programme was used for MetG PCR: 94 ºC for 10 mins; 34 cycles of 94 ºC for 30 sec, 53 ºC for 30 sec, 72 ºC for 90 sec and finally for 5 min at 72 ºC. RNA integrity was assessed and measured using Quant-iT^{TM} Ribogreen ® RNA Reagent and Kit.

5.5.1.5 First Strand cDNA Synthesis From Total RNA (Reverse Transcription)

The RevertAid ^{TM} First Strand cDNA Synthesis Kit (Cat No: K1621 and K1622) was used to synthesize cDNA from previously extracted and DNase I treated total RNA. The RNA used to synthesize cDNA was in the range of 10 ng - 5 µg. The components of the kits were thawed, mixed and briefly centrifuged. 10 µl of Total RNA (0.1-5 µg), 1 µl of random hexamer primer was added into a sterile, nuclease free tube on ice. The tubes were incubated at 65 ^{o}C for 5 min. The following components were added in the above mentioned order in a tube; 4 µl of 5 x Reaction Buffer, 1 µl of RiboLock^{TM} RNase inhibitor (20 U / µl), 2 µl of 10 mM dNTP Mix and 2 µl of M-MuLV Reverse Transcriptase (20 U/ µl) in a total volume of 20 µl. The components were mixed gently and centrifuged. The mixture was incubated for 5 min at 25 ^{o}C followed by 60 min at 45 ^{o}C and finally the reaction was terminated by heating at 70 ^{o}C for 5 min.

5.5.1.6 PCR Amplification Of the LNB Gene in cDNA

cDNA of *A. oris*-MG1 and *A. naeslundii*-NCTC 10301 was used to amplify the LNBP gene (Ana_1637) in order to confirm the presence or absence of mRNA. The reverse transcription reaction product was used directly in PCR using LNBP-F2, LNBP-R2 primers and run on the PCR programme (94 ^{o}C for 10 min, 45 cycles of 94 ^{o}C for 30 sec, 49 ^{o}C for 30 sec, 72 ^{o}C for 90 sec, and finally run at 72 ^{o}C for 5 mins). Products were run on a 1% agarose gel and visualized under UV light using gel imager.

5.5.1.7 Sequencing Of the LNB Amplicon

Prior to sequencing the PCR products of reverse transcription, fresh PCR product was run again on a gel and bands were visualized with the DarkReader ® transilluminator (Clare Chemicals Research, Dolores, CO, USA). The bands of interest were excised/cut with a sterile scalpel and collected in a sterile 1.5 ml centrifuge tube. To elute the DNA from the gel, the QIAquick Gel Extraction Kit (Cat No; 28704, Qiagen) was used and the protocol was followed as recommended by the manufacturer. The extracts were stored at -20 ^{o}C.

The amplicon was reamplified using the PCR conditions and primer pairs for LNBP as described above. The PCR products were run on agarose gel and subsequently cleaned with ExoSAP-IT® (Product No: 78200, Affymetrix, USA) using the manufacurer's protocol. The amplicon was sequenced using Big Dye terminator cycle sequencing kit using LNBP-F2 and LNBP-R2 primer set. The cleaned sequencing reaction products were run on an ABI sequencer 3730 x I (Applied Biosystems) and edited using BioEdit software and aligned with the reference sequence of LNBP amplicon downloaded from CMR database of *A. oris*-MG1.

5.5.2 Results

The RT-PCR experiment was conducted to demonstrate up-regulation of the LNB phosphorylase gene when *A. oris* (MG1) was grown in medium containing LNB. The electrophoresis image shown in Figure 5.18 has two blocks. The red block represents the experiment run with MetG gene as a control. The green block represents the actual experiment performed using the LNBP primers The following extracts were taken, D (genomic DNA), cG (cDNA grown in glucose), cO (cDNA grown in Oligosaccharides), RG (mRNA extract from the cells grown in glucose) and RO (mRNA extract from the cells grown in LNB). The control experiments with the MetG primers showed that the DNA extracted from *A. oris*-MG1 appeared positive and also the cDNA synthesized from the *A. oris*-MG1 cells grown in the presence of glucose and LNB appeared positive

while mRNA extracts isolated from cells grown in the presence of glucose and LNB showed no bands on the gel and this showed that there was no impurity present in the RNA when extracted using above mentioned protocol. The same extracts were used to run the RT-PCR experiment using the LNBP primer set (blue block). No bands were obtained from the RNA extracts grown in the presence of glucose and LNB. cDNA gave positive result when grown in the presence of LNB showing that the LNBP genes were expressed when grown in the presence of LNB. At the same time no band was observed for the cDNA from cells grown in the presence of glucose. Also, genomic DNA gave a positive band on the gel plate with the LNBP primers. Therefore the result indicates that Ana_1637 gene was up-regulated when *A. oris*-MG1 was grown with LNB and this result links the finding in the current study with the whole genome sequencing comparison that all *A. oris* strains have the LNB operon and suggests is better able to utilize LNB from mucin. The size of LNB amplicon obtained was 600 bp.

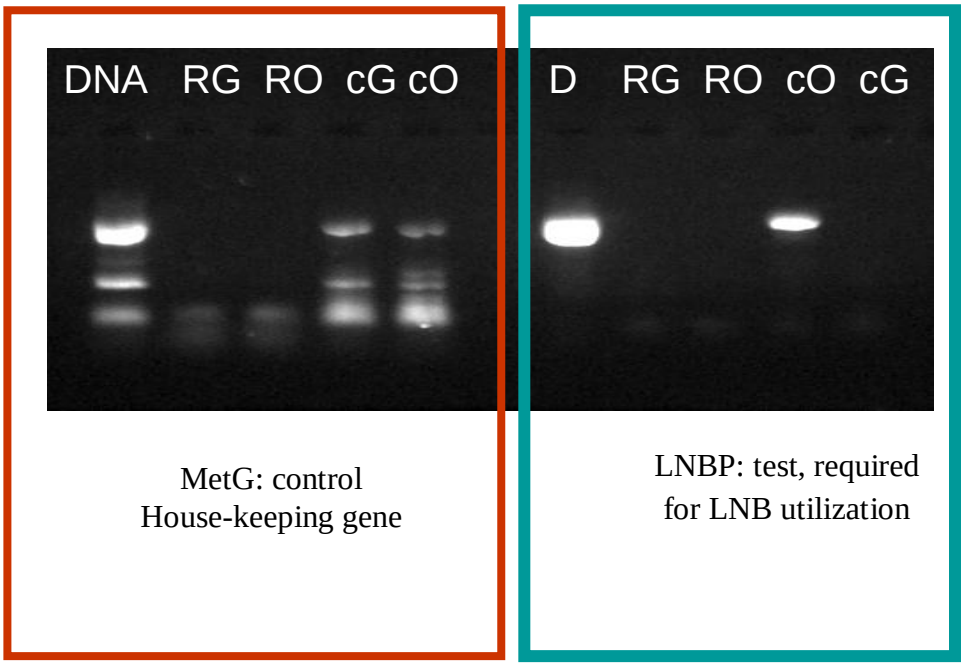


Figure 5.18: RT-PCR Demonstration of up-regulation of Ana_1637 genes

***Red block represents a control experiment using MetG primers and the blue block indicates RT-PCR test with LNBP primers. D (Genomic DNA), RG (RNA extract from glucose-grown cells), RO (RNA from LNB grown cells), cO (cDNA from LNB grown cells), cG (cDNA from glucose-grown cells)**

5.6 Growth Studies Of *Actinomyces* Spp. In The Presence Of Lacto-*N*-Biose I

5.6.1 Material And Methods

The *A. oris* and *A. naeslundii* isolates were characterized for their ability to grow in the presence of LNB oligosaccharides by using the following method. The strains are listed in Table 5.6.

5.6.1.1 ASS (Semi-Synthetic) Medium For Growing *Actinomyces* Spp.

The ASS basal medium was prepared with some modifications from the method of He **(2008)**. The basal medium was composed of **Amino Acids** L-glutamic acid Monosodium salt, Monohydrate, 2.54g/L, Sigma Aldrich, France; L-cysteine hydrochloride, anhydrous, 0.2g/L, Sigma, USA; L-leucine 0.15g/L, Fisher Scientific, New Jersey, USA; L-lysine Hydrochloride 0.15g/L, Calbiochem, USA; L-arginine Monohydrochloride, 0.21, Sigma, USA; L-proline, 0.006, Sigma Aldrich, St.Louis, MO, USA; L-glycine, 0.004, Fisher Scientific, New Jersey, USA; **Salt Solution I** K_2HPO_4 2.08g/L, BDH, AnalaR®; KH_2PO_4, 2.92 g/L, VWR International Ltd, Poole, England; NH_4CL, 0.54g/L, Sigma, St.Louis, MO, USA. **Urea Solution** Urea, 0.3g/L, Sigma Aldrich, USA. **Salt Solution II** $MgCl_2$ ($6H_2O$), 1.2g/L, Sigma, St.Louis, USA; $MnCl_2$ ($4H_2O$), 0.02g/L, ACROS organics, New jersey, USA. $CaCl_2$ ($2H_2O$), 0.11g/L, Sigma-Aldrich, USA; $FeSO_4$($7H_2O$), 0.02g/L, ACROS, Organics, New Jersey, USA. **Vitamin Stocks** Thiamine HCl, 1mg/L, Sigma Aldrich, St.Louis, USA; D-Biotin, 0.2mg/L, Sigma Aldrich, USA; Nicotinic acid, 2.0mg/L, Sigma, St.Louis, MO, USA; p-Amino-benzoic acid, 0.2mg/L, D-pantothenic acid-hemicalcium salt, 1.0mg/L, Sigma Aldrich, Germany; Pyridoxine monohydrochloride (vitamin B6), 2.0mg/L, Sigma, USA. Folic acid, 0.2mg/L, Sigma, USA; Riboflavin (vitamin B2), 0.1mg/L; Vitamin stocks were prepared separately, Filter sterilized and 0.3% Bacto™ Peptone, SLS, USA was added in the final medium (Add 0.2ml of 10% peptone into 6ml of medium).

5.6.1.2 PYE Medium For Growing *Actinomyces* spp.

The modified basal medium Peptone Yeast extract (PYE) was also used in this study to characterise the growth efficiency of *A. oris* isolates. PYE medium was composed of 10 g/l tryptone, 6 g/l yeast extract, 200 µl Tween 80, 0.5 g/l cysteine hydrochloride, 20 ml/l salt solution A, 20 ml/l salt solution B. Stock solution Salt A was made up by dissolving 0.16 g of anhydrous calcium chloride and 0.16 g magnesium sulphate in 400 ml distilled water. Salt B was prepared by dissolving 0.8 g di-potassium hydrogen orthophosphate tri-hydrate, 8.0 g of sodium hydrogen carbonate, 1.6 g of sodium chloride and 0.8 g of potassium di-hydrogen orthophosphate in 400 ml of distilled water. Both stock solutions were autoclaved at 121 ºC for 15 mins and stored at 4 ºC. The basal medium was adjusted to pH 7 with 1M hydrochloride acid. Sugars were added at a concentration of 1% (w/v) to the basal medium and autoclaved at 121 ºC for 15 minutes. Lacto-*N*-Biose I was filtered sterilised and added after autoclaving.

5.6.1.3 BASAL medium And Inoculum Preparation

The ASS medium was prepared and cold sterilized after adding 0.3 % peptone and similarly PYE medium was prepared. ASS and PYE medium was divided into three parts separately. In the first part of the medium 1% Lacto-*N*-Biose I (an oligosaccharide obtained from Dr. Motomitsu Kitaoka, Japan) was added, while in the second portion of the ASS and PYE medium 1% glucose was added and in the third portion of ASS and PYE medium distilled water was added in place of substrates as a control. 86 *Actinomyces* strains were selected for characterising growth which includes 44 strains of *Actinomyces oris* and 42 strains belonging to *Actinomyces naeslundii*. The names of these *Actinomyces* strains were shown in Table 5.6.

$$\text{Volume of sterile broth needed to dilute the bacterial suspension} = \frac{\text{Volume of bacterial suspension (OD of suspension} - \text{OD of blank} - 0.5)}{0.2}$$

The strains were grown on FAA plates for 48 hours. The inoculum was prepared as a homogeneous suspension in 500 µl of sterile ASS medium. The turbidity of cells was adjusted to 0.2 OD using the above formula and 10µl of bacterial suspension was added in each well of 96 well plates. The wells were marked for each strain on the plate. 190 µl of ASS medium containing lacto-N-biose oligosaccharides, ASS medium containing glucose and only ASS medium was added to the marked wells. The plate was left in anaerobic cabinet at 37 ºC to pre-warm the medium. The plate was sealed with clear adhesive film (MicroAmp^{TM}, USA) by keeping the plate inside the anaerobic incubator to maintain anaerobic conditions inside the 96 well plate and then put in iEMS Reader (Labsystems, iEMS Reader MF) to measure the OD. The Ascent software was used for reading optical densities. The programme was set up at step time for 15 mins, measurement mode selected was continuous, measurement type chosen was kinetic, and filters used at 620 nm and measurement counts were 400. Shaking conditions were also set for total time 10 sec, and started at 10 sec, and stopped at 20 mins and incubation temp was set up at 37 ºC. The pH was taken for each well using a pH meter (Oakton pH 5 Acorn series, Singapore).

Table 5.6: Strains of *Actinomyces* used in Lacto-*N*-Biose Growth experiment

Serial number	*A. naeslundii*	Serial number	*A. oris*
1	NCTC 10301	1	S62B
2	E1-20	2	F5A1
3	M42-1-1	3	M47-1-1
4	TM-3	4	F24C1
5	Pn16E	5	U149-1
6	R24330	6	F23A1
7	T20P-1	7	WE9A10-1
9	S49H	8	P5K
10	461 CCUG 34725	9	R14746
11	R9841	10	F7j1
12	R19039	11	A2A-1
13	S38H	12	F11C1
14	S41C	13	G140C
15	U136-1	14	MG1
16	R8152	15	R21091
17	F6E1	16	G53E
18	G126D	17	CCUG 33920
19	Pn1GA	18	MMRCO2-1
20	WE7B1	19	R23275
21	R13240	20	A7A-1
22	CCUG 35334	21	CCUG 34286
23	S65A	22	A19A-1
24	T9P-1	23	S24V
25	S44D	24	P6N
26	MB-1	25	WE8B-23
27	R24330	26	F4D-1
28	G127B	27	R23275
29	Pn6N	28	MMRCO6-1
30	R8152	29	A18A-3
31	R19039	30	F2A12
32	MMRC12-1	31	WE10B-1
33	T23P-1	32	G48B
34	S53C	33	S33A
35	CCUG 37599	34	UN4B6c
36	F12B1	35	R23277
37	T14P-1	36	M46-1-1
38	G51C	37	W11-1-1
39	F6E1	38	S64C
40	S43L	39	F28B-1
41	WE6B-3	40	M48-1B-1
42	W8-2-3	41	R11372
		42	S55N
		43	A7A-1
		44	F21A1

5.6.2 Results

5.6.2.1 Growth Results In ASS medium With And Without LNB

The results of growth experiments were obtained using a modified basal medium (ASS & PYE) specifically designed for the growth of *Actinomyces* to observe the difference of growth in *A. oris* and *A. naeslundii* spp. with added LNB and without LNB. The 44 strains of *A. oris* selected (Table: 5.6) gave a mean final optical density (OD) of 0.213 ± 0.07 in ASS medium containing 1% LNB while the mean final pH of *A. oris* strains observed was 4.99 ± 0.36. Similarly the growth results obtained for *A. naeslundii* strains grown in ASS medium containing 1% LNB oligosaccharide were mean final OD of 0.089 ± 0.05 and a mean final pH of 6.28 ± 0.18 (Table 5.7)

Table 5.7: Summary table for Growth of *Actinomyces* in ASS Medium + 1% LNB

	Mean Final OD	**Mean Final pH**
***A. oris* (n=42)**	0.213 ± (0.07)	4.99 ± (0.36)
***A. naeslundii* (n=44)**	0.089 ± (0.05)	6.28 ± (0.18)

5.6.2.2 Growth Results in PYE Medium Containing LNB And Without LNB

Another basal medium PYE containing 1% LNB was used to observe the growth pattern of *A. oris* as compared to *A. naeslundii* strains. The mean OD observed for *A. oris* strains in PYE medium containing 1% LNB was 0.167 ± 0.07 and the mean pH was 5.53 ± 0.31 while the OD observed for *A. naeslundii* strains was 0.069 ± 0.02 and the mean pH was 6.23 ± 0.21 (Table 5.8). The results showed that *A. oris* grew better on LNB than *A. naeslundii* and the OD increase and decrease in pH were significantly greater for the *A. oris* strains than for the *A. naeslundii strains* as shown by t-test ($p \leq 0.001$). This may be of ecological significance. There was no significant difference found between the two species in their growth values obtained using the other two media. ASS medium proved to be a better basal medium because the *A. oris* growth was more profound.

Table 5.8: Summary table for Growth of *Actinomyces* in PYE Medium + 1% LNB

	Mean Final OD ± SD	Mean Final pH ± SD
***A. oris* (n=42)**	0.167 ± (0.07)	5.53 ± (0.31)
***A. naeslundii* (n=44)**	0.069 ± (0.02)	6.23 ± (0.21)

5.7. Generation Of Knock-Out Mutants (Δ ANA_1637) Of *A. oris*-MG1

5.7.1 Introduction

A. oris is a Gram positive and facultative anaerobic micro-organism. *A. oris* is commonly isolated from the human oral cavity and is actively involved in plaque formation (Hintao *et al.*, 2007; Kamma *et al.*, 2000; Papaioannou *et al.*, 2009; Preza *et al.*, 2008; Preza *et al.*, 2009). *Actinomyces* species are involved in the initial colonization of the tooth surface which in turn provides an adhesive platform for the colonization of bridging bacteria which finally attract late colonizers of bacteria to form a biofilm on the tooth surface (Kolenbrander *et al.*, 2006; Ritz, 1967). There were in vitro studies carried out to see the adhesion properties of *A. oris* with *F. nucleatum* (Periasamy *et al.*, 2009). *A. oris* was shown to have cell surface fimbriae. Type I fimbriae promote the initial attachment of *A. oris* to host salivary proline-rich proteins (Gibbons *et al.*, 1988), while type II fimbriae are responsible for lectin-like activity which mediate host cell activation and biofilm development (Cisar *et al.*, 1979; Costello *et al.*, 1979; McIntire *et al.*, 1978). It was hypothesized that binding of specific oligosaccharide motifs (GalNAcβ1-3Gal) was involved with receptor polysaccharides of early colonizing bacteria (Cisar *et al.*, 1995; Cisar *et al.*, 1997) or with host cells (Ruhl *et al.*, 1996; Ruhl *et al.*, 2000; Stromberg & Karlsson, 1990). The host cell receptors were also masked by terminal α2-3 linked sialic acid, therefore for type 2 fimbriae-mediated attachment of cells pre-treatment by sialidase enzyme was required. Sialidase is an enzyme which is produced by *A. oris* (Mishra *et al.*, 2010). The recent study also reports that *A. oris* has the ability to utilize LNB type I oligosaccharides from mucins which may mediate the formation of oral biofilm. To dissect the molecular function of these genes and their specific role, there was a need to engineer specific LNBP gene deletion mutants of *A. oris* using state-of-the art gene manipulation technology. The allelic exchange in *Actinomyces oris* (MG1) was nearly impossible in the past due to the lack of facile gene disruption technology. Traditional methods were used for genetic manipulation in *A. oris* which had some drawbacks. The old strategy adapted for allelic exchange was with using nonreplicative plasmids as delivery vectors in *A. oris* (Chen *et al.*, 2007; Yeung & Ragsdale, 1997; Yeung *et al.*, 1998). This technique produces polar

mutations which affect the downstream genes. This problem was handled successfully by developing a method which generates non-polar, in-frame deletion of the sortase gene in 2007 (Mishra *et al.*, 2007). Subsequently, our studies of deletion mutant (Ana_1637) reported to use the novel approach of allelic exchange technology which enabled us to generate unmarked in-frame deletion mutants of *A. oris* in less time and efficiently. Subsequently, our studies of Ana_1637 deletion mutant of *A. oris* have led to the surprising discovery that the LNBP gene is solely responsible for the utilization of type I oligosaccharides in mucins. These findings in conjunction with genome sequencing opens a molecular approach to understanding the role of LNBP gene in biofilm formation and exploring possible novel strategies for the improved control of dental plaque formation.

5.7.2 Development Of An In-Frame Deletion System Based On galK For *Actinomyces* species.

GalK feasibility as a counter selectable marker in *A. oris* was already investigated in detail (Mishra *et al.*, 2010) and therefore this strategy was employed in our study for preparing deletion mutants. The mutant was prepared in a two step process. In the first step the suicide plasmid was integrated at the target locus (Ana_1637) by homologous recombination. During the second step the plasmid was excised via a second cross over event that leaves an unmarked deletion of the target gene. *A. oris* has a galK homologue (gene ID ANA_0203, *A. oris* MG1 genome database from http://www.oralgen.lanl.gov). Galactokinase (galK) is a key enzyme in galactose metabolism and phosphorylates galactose to yield galactose-1-phosphate. GalK functional ability in MG1 was tested with its sensitivity to 2-deoxygalactose (2-DG). This enzyme converts 2-DG into 2-deoxygalactose-1-phosphate, which accumulates in the cell and is toxic (Reyrat *et al.*, 1998). Therefore when cells of the MG1 [wild type (WT)] were spread on the plates containing 0.25% 2-DG, no colonies appeared. This indicates that *A. oris* galK gene might act as a counter selection. This suicide effect has been exploited in a negative selection procedure in some of the bacteria previously (Merritt *et al.*, 2007; Warming *et*

al., 2005) and is used in the present study. The counter selection strategy for generating the in-frame deletion experiment is explained in Figure 5.19

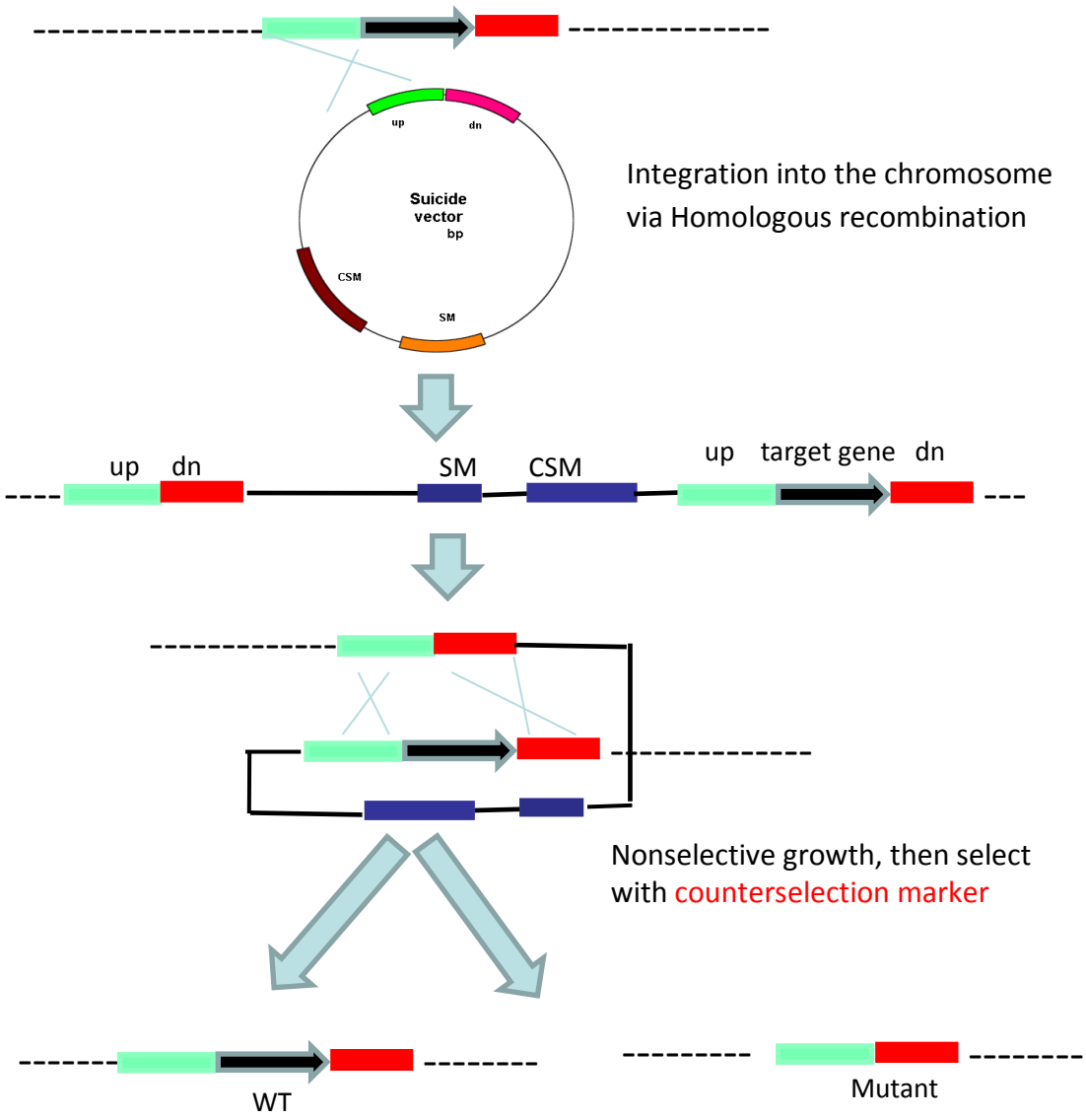


Figure 5.19: Counter selection strategy.

*The suicide vector required both a positive and a negative selection mechanism; After the suicide vector has integrated onto the chromosome, then applied the negative selection pressure; Therefore, only strains that have undergone the second recombination to cut the plasmid out will survive. In this case, the ratio of wild-type to in-frame deletions mutants should be 5:5.

**SM: Selection marker ; CSM: Counter Selection marker ; dn: down ; WT: wild type.

5.7.3 Material And Methods

The mutant of MG1 (Ana-1637) was generated to test the effect of LNB oligosaccharides on LNB operon expression when grown in the presence of LNB oligosaccharides.

5.7.3.1 Generation Of In-Frame Deletion Mutant Of *A. oris*-MG1 Lacking Gene Ana_1637

The protocol used to generate an in frame deletion mutant was adapted personally by visiting the Department of Microbiology and Molecular genetics, University of Texas-Houston Medical School under the supervision of Hung Ton-That and also using their published papers (Mishra *et al.*, 2010; Wu & Ton-That, 2010). To create an in-frame deletion construct of *Ana*_1637, a 2.0 kb fragment from the upstream and downstream of *Ana*_1637 was amplified by PCR using primers 1637upF(XbaI) and 1637upR(EcoRI) or 1637dnF(EcoRI) and 1637dnR(kpnI) respectively. The appropriate restriction sites were incorporated into 5' and 3' of primers (Table 5.9).

Table 5.9: PCR primers for gene (Ana_1637) mutant generation of *A. oris*-MG1

1637upF(XbaI)	*ggcg tctaga**GTTCAGGGGA CTTCAGGCCT TGACA
1637upR(EcoRI)	*ggcg gaattc**CGGCGCCGCG GGCTGGGTCT GGAGAG
1637dnF(EcoRI)	*ggcg gaattc**ACCAGTGGCT CACCGCTGCC TCGCAG
1637dnR(kpnI)	*ggcg ggtacc**AACAGCAGGA CCCGCCCCAG TCCCT

* Restriction sites in small letters
** Appropriate gene sequence of *Actinomyces oris*-MG1

5.7.3.2 Polymerase Chain Reaction

The reagents used for PCR were 5x Phusion HF Reaction Buffer, B0518S, NEB,100µl; dNTPs ,1µl; Phusion High Fidelity DNA polymerase, M0530L ,0.6-1.0 µl; 1.5 µl of 1637 Forward Primer; 1.5 µl of 1637 Reverse Primer and 1.5 µl of DNA template. PCR was set up using 1637up (Xbal EcoR1) and 1637dn (EcoR1 Kpn1) primers. PCR products were purified using the gel purification method. Then the PCR products were run on a gel and bands were cut under white light to save the bands from damage by UV light. The cut bands were kept in eppendorf tubes and 600 µl of solubilization buffer QG (QIAquick Gel Extraction Kit) was added and finally the tubes were kept in water bath for 6 mins and kept on mixing after every two mins until the gel dissolved. Then 600µl of solution was loaded on to columns and spun at 13,000 rpm for 1 min. 500µl of QG buffer was added to the column and was spun. The contents were washed with PE buffer and centrifuged for 1 min. Finally the mixture was spun for 2 mins to evaporate the residual ethanol and was eluted in 40 µl of sterilized water.

5.7.3.3 Digest

The above PCR products were digested for upstream 1637 up (Xbal EcoR1) and downstream fragment 1637 dn (EcoR1 Kpn1) DNA was cleaved with two restriction enzymes simultaneously to save time. The website (www.neb.com/nebecomm/doubledigestcalculator.asp) was used to find appropriate NEB buffer (NEBuffer4 + BSA ; NEBuffer4 was used in this case). PCWU2 vector (Figure 5.20) and PCR products were digested with appropriate digestion mix.

Then 10 µl of pre-cut plasmid PCWU2 (Deletion vector/suicide plasmids) + 70µl of H_2O + 7µl of NEBuffer4 + 3µl of BSA + 5µl of EcoR1 + 5µl Xbal were taken and mixed gently and finally spun for a short time. The products were left overnight or for 3-4 hours in 37 ºC incubator. Finally the purification was performed using GenElute PCR cleanup Kit (Cat No: NA 10201-1KT; Sigma) and concentration was measured using a Nanodrop fluorometer.

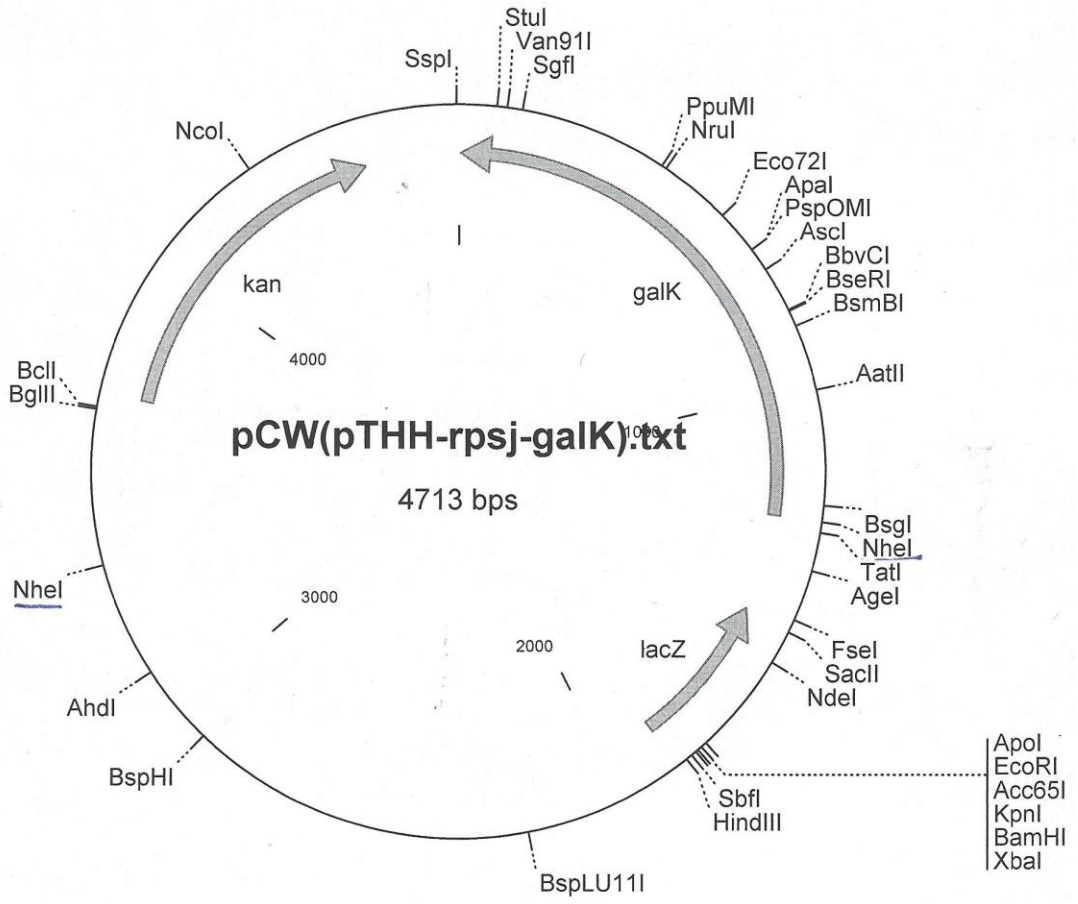


Figure 5.20: The restriction sites on suicide/deletion vector PCWU2.

5.7.3.4 Ligation

The digested products were ligated into pre-cut plasmid PCWU2 (Figure: 5.23). The Quick ligase was used (Cat No: M2200S of NEB). 10 µl of Quick Ligation Buffer, 3µl upstream and 3µl downstream, 2µl of plasmid PCWU2 and 2µl of H_2O was added and contents were mixed gently and left on ice for 10 mins.

The ligation mix was incorporated gently into competence cells of *E.coli* and was left on ice for 30 mins. Heat shock treatment was given for 1 min by putting them in water bath and then left on ice for 2 mins. 1ml of LB broth was placed in eppendorf tubes and transformants were added. The contents were left on a rotary shaker at 37 ºC for one hour. They were spun at 3,000 rpm for 1 min. Supernatants were taken off leaving behind the 100 µl of supernatant only. The contents were mixed gently with 100 µl of pellet in tube and were restreaked on plates. The plates were incubated at 37 ºC.

Colonies containing the integrated plasmid were picked (1637-PCWU2) and grown in LB broth. 1.5 µl of overnight culture was placed in eppendorf tubes and centrifuged for 1 min. The cells were dispersed in 250 µl of P1 buffer and 250 µl of P2 was added. 400 µl of neutralization buffer was added and it was centrifuged for 10 mins. Supernatants were poured on columns and tubes were centrifuged. Cells were washed with PE buffer and eluted in 45 µl of water.

5.7.3.5 Transformation

The plasmids generated from the above procedure were transformed into CW1 strain (Δ galK) by electroporation. Competent cells of CW1 were prepared well ahead of time.

5.7.3.6 Competent Cells Of CW1 (Δ galK)

The competent cells of *A. oris* CW1 (Δ galK) were the host for carrying galK. Fresh clones were picked out into 6ml of heart infusion broth (HI) for overnight growth. The 6ml of overnight culture was poured into 60ml of fresh HI broth in 125 ml flask and incubated until the OD 0.5 to 0.6 was reached (about 5 hour growth). Then freshly prepared 35ml of 15% glycine (Glycine was dissolved in HI broth and filter sterilized) was added. The cells were incubated for 1h at 37 ºC in a water bath shaker and then the cells were collected by centrifugation at 8,000 rpm for 6 mins. The cells were washed twice with 30 ml of 10% ice-cold glycerol. 0.8ml of the 10% glycerol was re-suspended and aliquots 180 µl, kept frozen in an ethanol-dry ice bath. The competent cells of *A. oris* CW1 (Δ galK) were stored at -80 ºC.

5.7.3.7 Electroporation

The apparatus used for electroporation was BIO-RAD (PC module, CE Module, Gene PulserXcell). The program was set on Gene pulser (Capacitance µF = 25, Persistence = 400Ω, Cuvette = 0.2cm and Voltage 2.5kv). The special cuvettes were used called MBP cat no. 5520 (Molecular BioProducts, Electroporation cuvettes, Pre-sterilized). Pre chill the 0.22 µm Biorad gene pulser cuvettes at -20 ºC and also pre-warm the HI broth in 37 ºC. The competent cells were thawed on ice for about 15 mins and also the plasmids were thawed on ice. 15 µl of plasmids was taken and added to 150 µl of competent cells and the entire contents transferred into bottom of chilled cuvettes. The cap was replaced and contents were gently tapped and were put on a gene pulser. The cells were transferred to pre-warmed growth medium. The cells were incubated for 1h to 3h in order for the cells to recover. The tubes were centrifuged and plated on HIA with 50 µg/ml kanamycin in it. This is for a single cross over event. The plates were left for 3 days of growth, some clones were seen. 2-3 clones were picked and restreaked on HIA with kanamycin. The 2 clones were inoculated into HI broth without kanamycin, they were grown till stationary phase (called second cross over) as shown in Figure 5.21.

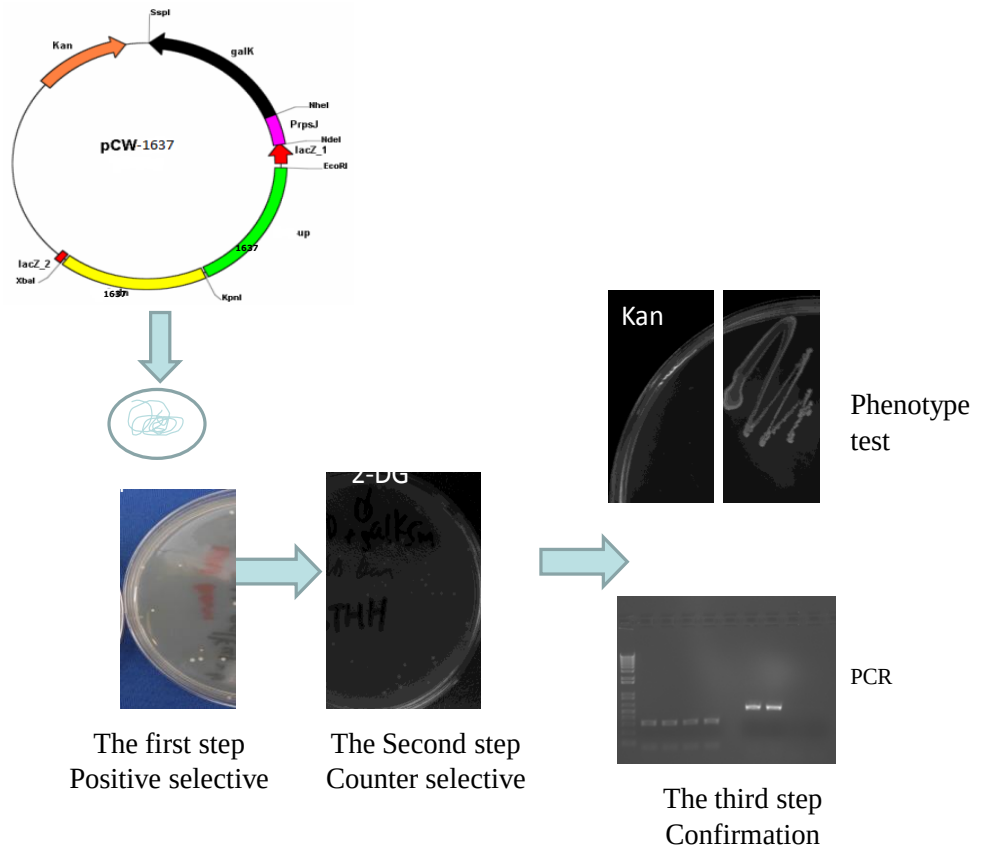


Figure 5.21: Procedure of generating an in-frame deletion mutant of Ana_1637

The dilutions were made 100x and plated 100 µl of diluted culture onto HIA with 0.25% 2-DG (4g of 2DG was taken and dissolved in 100 ml of H2O and marked as stock solution. 1.3 ml of stock was taken and added in agar plate and was left for 3 days.

After 3 days growth, 10-20 clones were randomly picked and restreaked onto HIA plate and HIA plate with kanamycin, the later will further confirm the clone lost the plasmid. PCR was done to check the wild type and mutant clones.

5.7.3.8 Plasmid Construction (Complementation Of The Deletion Mutant With pJRD_1637).

The intergenic region between Ana_1636 and Ana_1637 is very short, therefore it was speculated that the two genes consist of an operon and share the same promoter. Thus, the promoter region of the Ana-1636 was generated by PCR with the primer pair Pro-1636-F/R and the Ana_1637 fragment was amplified with primers com-1637-F/R (Table 5.10). The Ana_1636 promoter amplicon was digested with KpnI and XbaI and the Ana_1636 amplicon was digested with XbaI and EcoRI. The two fragments were then ligated into the KpnI and EcoRI sites of *E. coli*/*Actinomyces* shuttle vector pJRD215 to generate pJRD_1637. PJRD215 (shuttle vector) was bigger in size and contained Kanamycin and Streptomycin resistance cassettes. MG1 also has the same antibiotic resistance cassettes on its genome. The Complement vector is generated to observe that the mutation was restored to the wild type phenotype and has not lost its growth characteristics lacto-N-biose I.

Table 5.10: Primers for Plasmid construction

Pro-1636F(kpnI)	ggcg ggtacc CATGCTTCCG GCCGTGATGG GCAGCA
Pro-1636R(xbaI)	ggcg tctagaCGTCGGTGTC CCGGTTGGTC CTGCGG
com-1637F(xbaI)	ggcg tctagaCCGCCACCAG GCATATCGAC CGCCT
com-1637R(EcoRI)	ggcg gaattcTCACTGCGAG GCAGCGGTGA GCCAC

5.7.4 Results

5.7.4.1 *A. oris* ΔgalK (CW1) Generation As A Host Strain

A mutant of GalK (ΔgalK) and kanamycin sensitive colonies were prepared in a previous study by Mishra (Mishra *et al.*, 2010) and these were used in our experiment. These colonies were prepared using a procedure described in the published paper by Mishra. Wild type *A. oris*-MG1 is kanamycin sensitive and 2-DG sensitive. The plasmid containing the mutant of galK was constructed and which contains galK but is missing the bulk of the galK cistron and so named as pCW-ΔgalK. The plasmid was transformed into *A. oris*-MG1 and called as integrant which, was kanamycin resistant and 2-DG sensitive. The galK deletion mutant was obtained by growing an above mentioned kanamycin resistant colony in the absence of antibiotic and plating on BHI plates containing 0.25% 2-DG. 12 colonies were picked and tested for their kan-sensitivity and galK deletion gene (ΔgalK) by PCR. Finally when these plasmid expressing galK were introduced into ΔgalK mutants then these bacteria will have lost the ability to grow in the presence of 2-DG. Therefore the ΔgalK mutant strain and the GalK-expressing plasmid together can make a powerful system to generate an in frame deletion mutant of *A. oris* which can subsequently be used to create specific in-frame deletion mutants in the Ana_1637 gene or LNBP operon.

5.7.4.2 Generation Of An In-Frame Deletion Mutant Of *A. oris*-MG1 (Ana_1637)

A kanamycin resistant colony was grown in HIB without antibiotic overnight as shown in Figure 5.22 (a). The following day, cultures were diluted 100- fold and 100 µl of the diluted culture was plated on BHI agar plates supplemented with 0.25% 2-DG. These plates were incubated for 4 days until 2-DG-resistant colonies appeared as shown in Figure 5.22 (b).

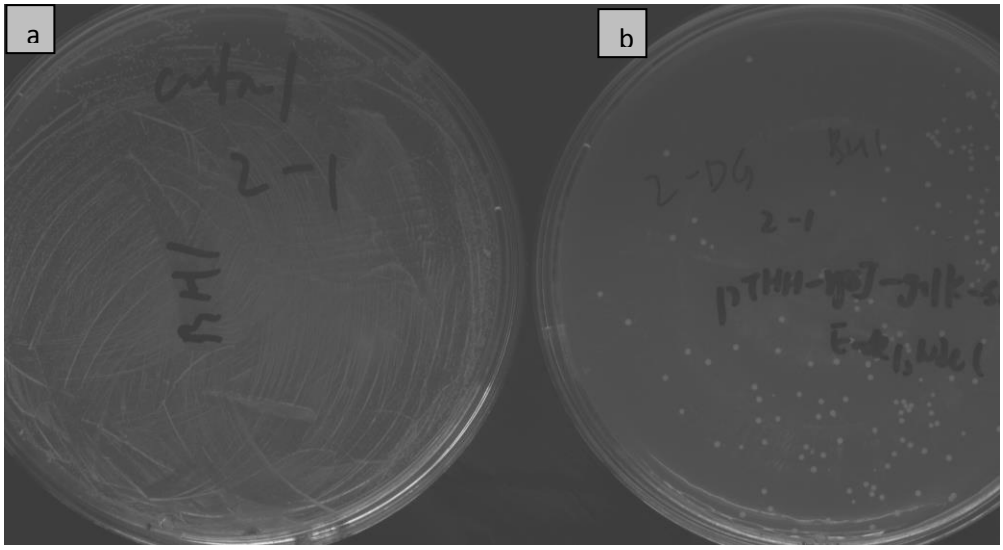


Figure 5.22: Growth observation on plates containing antibiotic and 2-DG

(a & b: a. Antibiotic –resistant clones purified and grown in BHI medium overnight without antibiotic. b. 2-DG resistant colonies appeared on plates containing BHI medium supplemented with 0.25% 2-DG).

Then 12 randomly chosen 2-DG-resistant isolates were restreaked on BHI containing 2-DG plates and Kan plates at the same time, to confirm to be kanamycin-sensitive as shown in Figure 5.23 (a and b). Kanamycin-sensitive colonies were inoculated in BHI for overnight growth and finally genomic DNA was extracted. 2-DG-resistant colonies underwent the second recombination event, which would either complete the allelic exchange or form WT.

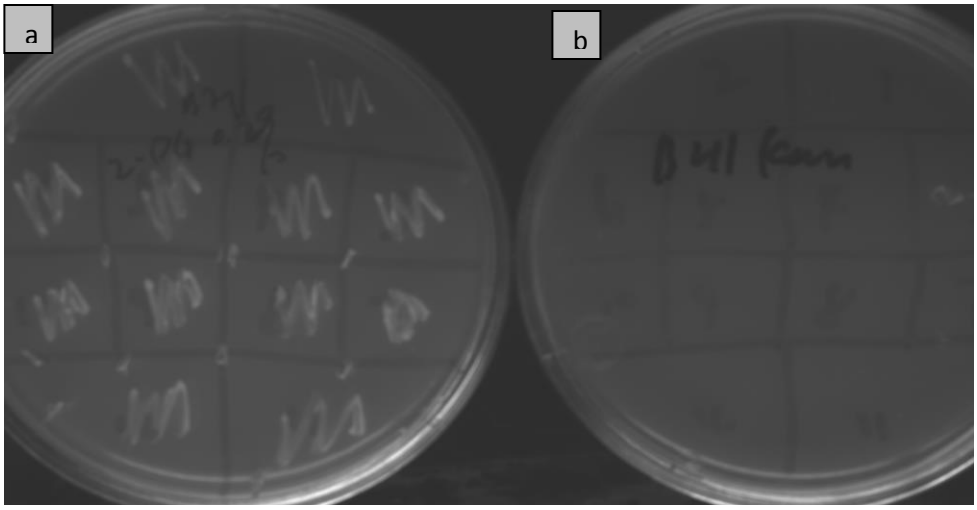


Figure 5.23: Growth confirmation of mutant colonies on plates

(a, 2-DG resistant colonies were streaked on BHI medium supplemented with 2-DG. b, 2-DG resistant clones were poured on BHI medium supplemented with kanamycin).

Polymerase chain reaction was used to determine the genotype of each kanamycin-sensitive isolate using the primers 1637-up-F and Down-R (Figure 5.24). LNB phosphorylase knockout mutants were obtained with 50% efficieny. 20 knockout mutant was complimented and recombinant clones were obtained

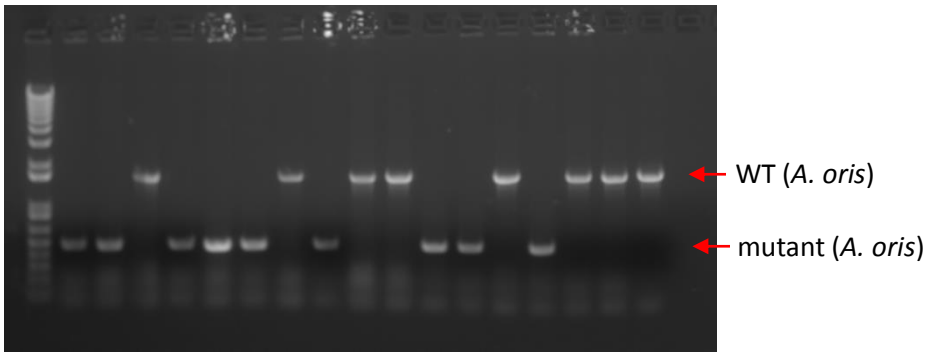


Figure 5.24: PCR demonstration to determine the genotype

*** The genotype of each antibiotic-sensitive isolate was determined using the primers 1637-up-F and 1637-dn-R**

5.7.4.3 *A. oris*-MG1 (WT) And The Mutant (Δ1637) Growth

5.7.4.3.1 Material And Methods

The growth studies were carried out to monitor the difference in growth of *A. oris*- MG1 wild type, *A. oris*-MG1 mutant (Δ1637) and complemented strains in ASS medium supplemented with 0.5% glucose, 0.5% lacto-*N*-biose and only in basal medium. For this purpose the stationary phase cultures of wild type and mutant strains (Δ1637) of MG1 were grown overnight in heart infusion broth (HI; Difco) at 37 ºC in a shaking waterbath. The overnight cultures were pelleted by centrifugation and washed twice with ASS medium, then the starter culture was normalized to an OD of 1.0 at 600nm and was inoculated at 1:50 dilution into ASS medium. 200 µl of this cell suspension was added to each well of a 96 well plate, and cell growth was monitored by a microplate reader (Tecan Infinite M100) at 600 nm. The OD_{600} value for each strain at a given time point was calculated as the average of triplicate OD_{600} values. The values read with a Tecan infinite M100 spectrophotometer were converted into corresponding values of a normal spectrophotometer.

5.7.4.3.2 Results

The results are shown in Figures 5.25 and 5.26. ASS medium only did not support the growth of WT, ANA_1637 mutant or its complemented strain. There was no significant difference in growth observed among the three strains when 0.5% glucose in the medium was added. ASS medium was able to differentiate the growth of *A. oris*-WT versus Ana_1637 mutant when supplemented with Lacto-*N*-biose. The OD of *A. oris*-WT was observed to be 0.7 after 40 h incubation, while the mutant strain lacking the Ana_1637 gene can only show growth to an OD of 0.3 in the same duration. As expected, introduction of a WT-copy of Ana_1637 in the Ana_1637 mutant strain restored its growth, albeit at a lower OD value compared with WT. The reason may be the complementing plasmid could not fully complement *A. oris*-WT level probably was due to the promoter region of Ana_1637, which was chosen to design the construct rather than as a complete promoter.

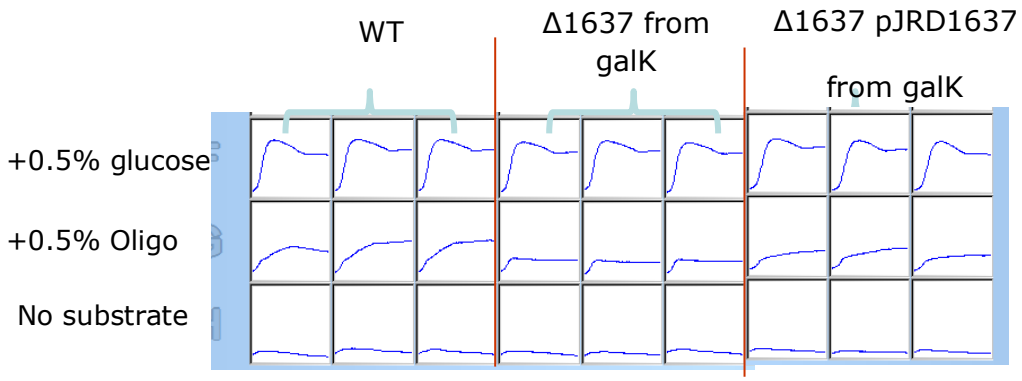


Figure 5.25: Growth of MG1 (WT), Δ1637 from galK, Δ1637 pJRD1637 from galK

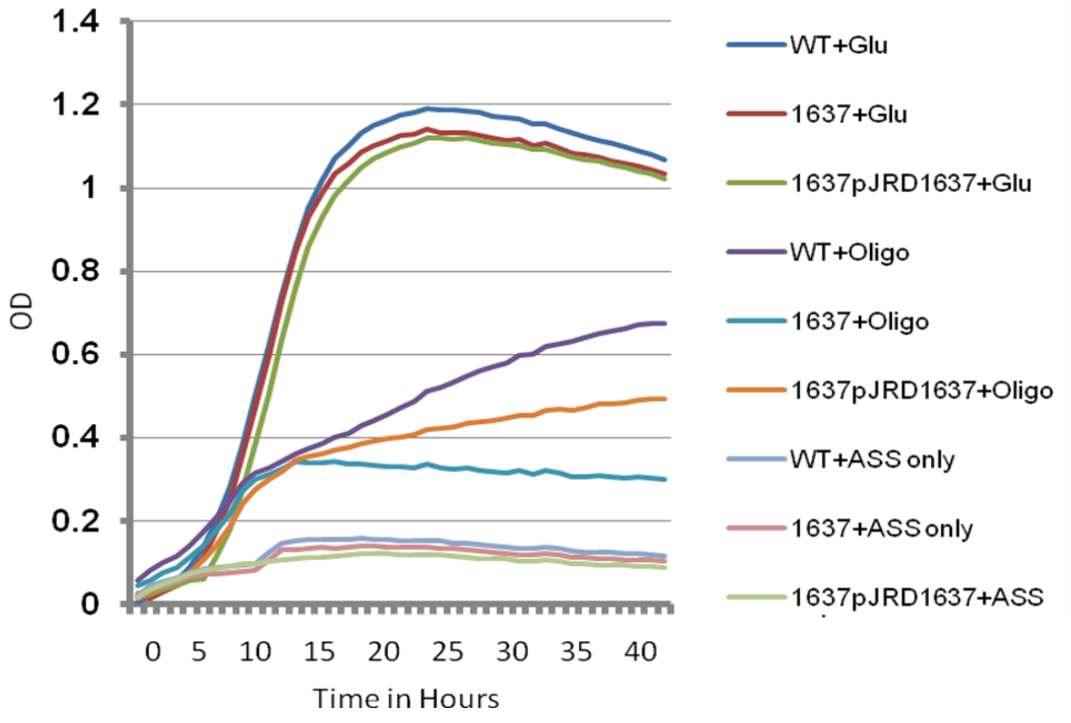


Figure 5.26: Wild MG1, Mutant (Deleted Ana_1637 gene) MG1 and Complemented (ANA_1637pJRD1637)

*The cells were grown in ASS medium containing Glucose (Highest peaks), Oligosaccharides containing 1% LNB (Medium peaks) and only in ASS medium

5.8 Discussion

In the current study it is reported for the first time that strains of *A. oris* group possess a Lacto-*N*-Biose metabolic pathway similar to that which was reported in various studies of the genomes of *Bifidobacterium*. This enzyme was identified for the first time in *Bifidobacterium bifidum* (Derensy-Dron *et al.*, 1999). Bifidobacteria are Gram-positive, intestinal, anaerobic bacteria found to colonize breast fed infants (Nishimoto & Kitaoka, 2007a; Rotimi & Duerden, 1981) while such colonization of bifidobacteria is not observed in bottle fed infants (Bezkorovainy, 1989). LNB was considered as a "bifidus factor" in human milk (Kitaoka *et al.*, 2005). Recently the *B. longum* subsp. *infantis* strain ATCC15697^T full genome was sequenced having a single circular chromosome of size 2.8Mbps (Sela *et al.*, 2008). The LNB metabolic pathway was revealed in detail in the bifidobacteria. The pathway starts from the activity of lacto-*N*-biose phosphorylase (EC 2.4.1.211, Blon_2174) with its transporter genes, an enzyme reported to be essential for the cleavage of Human milk oligosaccharides (HMO) (Nishimoto & Kitaoka, 2007b). The bifidobacteria are reported to contain seven genes responsible for LNB utilization and these genes were conserved among *B. longum subsp infantis* but were not present in *B. adolescentis*, and therefore the later is not capable of utilizing milk oligosaccharides structures (Ward *et al.*, 2006). Several bifidobacterial strains (including some infant associated ones) can not utilize milk oligosaccharides (LoCascio *et al.*, 2007; Ward *et al.*, 2006). Similarly the galacto-N-biose phosphorylase (GNBP) gene was discovered from the genome of *Clostridium perfringens* ATCC13124, a Gram-positive anaerobic intestinal bacterium (Nakajima *et al.*, 2008). The GNBP gene sequence in this bacterium is homologus to the Lacto-N-biose phosphorylase (LNBP) gene identified from various bifidobacteria. The colonization of *C. perfringens* in the intestine was justified by the ability of a bacterium to metabolize GNB which is an integral part of mucin core sugars. The LNBP was also identified from *Vibrio vulnificus* CMCP6, a Gram-negative gammaproteobacterium, often found in seawater, fish, plankton, and shellfish (Randa *et al.*, 2004). The genome of this species was sequenced recently (Kim *et al.*, 2003) because of its infection-causing ability in patients of cirrhosis, diabetes, or immunodeficiency syndrome. The presence of homologus genes to LNBP was investigated. Lacto-*N*-Biose (LNB) is a building unit of three type I

human milk oligosaccharides (HMO) lacto-N-tetraose, lacto-N-fucopentaose I, and lacto-N-difucohexaose I. Recently it has been found that *Bifidobacterium bifidum* is highly adapted for utilizing human milk oligosaccharides with a type I chain due to the presence of the LNB operon (Wada *et al.*, 2008). This organism is present in the gut of infants fed on breast milk due to its ability to utilize oligosaccharides from human milk (Sela *et al.*, 2008). In the present study the gene-by-gene sequence comparisons also indicate the presence of the LNB operon in *A. oris* strains but not in *A. naeslundii* strains. Such a difference may be of ecological significance since *A. oris* is more numerous in the mouth than *A. naeslundii* which may be at least partly due to the presence of the critical enzyme which is very helpful in degradation of type I glycans. The role of LNBP in stimulating bifidobacterial growth can be elucidated if the origin of LNB and GNB in humans is known. It was supposed that both GNB and LNB are structural parts of mucin and human milk respectively.

The Lacto-*N*-Biose was prepared practically in the lab in 2007 (Nishimoto & Kitaoka, 2007a) to understand the role of LNB as a "bifidus factor" in human milk. Sucrose and GlcNAc were used as a starting material. Four enzymes took part in the pathway e.g. sucrose phosphorylase (SP, EC 2.4.1.7), UDP-glucose--hexose-1-phosphate uridyltransferase (GalT, EC 2.7.7.12), UDP-glucose 4-epimerase (GalE, EC 5.1.3.2), and LNBP. 10-litre of reaction mixture having 660mM of sucrose and 600mM of GlcNAc along with four above mentioned enzymes were used and the yield was increased to 83% in the presence of UDP-Glc and Phosphate indicating that both act as a catalyst in the reaction. 1.4 kg of LNB of 99.6% purity was obtained after recrystallization treatment. Thereafter the role of these enzymes was observed in *Bifidobacterium longum* JCM1217 (Nishimoto & Kitaoka, 2007b). The activity of these enzymes suggest that all together they form an operon which is involved in lacto-N-biose I/galacto-N-biose metabolic pathway which was the cause of intestinal colonization of bifidobacteria in breast fed infants where they utilize lacto-N-biose I from HMO and similarly galacto-N-biose I from mucin sugars.

5.8.1 Elucidation Of The LNB/GNB Pathway

*lnp*A , *lnp*B, *lnp*C and *lnp*D are the enzymes encoded by the gene cluster responsible for LNB and GNB utilization/metabolic pathway. The structure of HMO containing LNB, which was cleaved with the help of enzymes called lacto-*N*-biosidases, not found in bifidobacteria. GNB is structural unit of mucin and was found to be cleaved with the help of enzyme called 'endo-α-*N*-acetylgalactosaminidase' was first reported in *Bifidobacterium longum* (Fujita *et al.*, 2005). LNB and GNB are believed to be carried into the cell due to the concerted action of three ABC transporter genes located upstream of the cluster (BL1638-BL1640). During the next step the LNBP acts to phosphorylate the GNB and LNB to form Gal1P and GlcNAc. Gal1P was converted into Glc1P by an enzyme GalE and then Glc1P was taken up by glycolytic pathway while GlcNAc/GalNAc was phosphorylated with the help of NahK to form GlcNAc1P/GalNAc1P which in turn enters into the aminosugar metabolic pathway. Operon encodes enzymes to metabolize LNB and GNB completely. This pathway is known as the LNB/GNB metabolic pathway (Fujita *et al.*, 2005). Similarly in the current study, the presense of these enzymes has been investigated in *A. oris* and these enzymes help to utilize the LNB-I from mucins using the LNB metabolic pathway. The pathway is demonstrated in figure 5.27.

5.8.2 Resemblance With Leloir Pathway

The Leloir pathway was observed in most organisms to metabolize Gal to Glc1*P* (Bouffard *et al.*, 1994; Frey, 1996; Leloir, 1951). Four enzymes take part in this pathway e.g galactose mutarotase (GalM), galactokinase, GalT, and GalE. LNB/GNB pathway is very similar to the Leloir pathway but there are some differences in the two pathways.

Table 5.11: Comparison of two metabolic pathways

	Leoir Pathway	**LNB/GNB pathway**
1	α-Gal is phosphorylated by Galactokinase with the help of one molecule of ATP	LNBP produces Gal1*P* without using ATP by the phosphorylysis of Galactosides
2	Same as above step uses ATP by GalK	GalNAc is transformed into GlcNAc1P with the ATP consumption using NahK.
3	The full set of enzymes encoded by gene involves *gal*M, *gal*K, *gal*T and galE along with LnpABCD cluster	*B. longum* contains only *gal*K and *gal*T cluster found in most of the organisms

For example, comparison of LNB/GNB pathway in *A. oris* as compared to *B. longum* and other species also showed some differences despite there being much similarity (Table 5.11). These differences were the absence of some genes encoding for the enzymes *gal*M and *gal*E in *A. oris* in the LNB/GNB pathway while LNBP and transporter gene presence is reported which utilizes LNB-I and this may be the main carbon or energy source in *A. oris* group of species and may describe the abundance of *A. oris* in the mouth as revealed by many studies.

In conclusion it was observed that the LNB/GNB metabolic pathway is the main pathway for galactose as a carbon or energy source in *B. longum*. Mucin carbohydrates and milk oligosaccharides may be metabolized using this pathway and may be the main reason for the "bifidus factor" for intestinal colonization by *Bifidobacterium* and the same was observed in the current study for *A. oris*, which utilizes oligosaccharides from mucins as their main carbon source.

5.8.3 Dose Dependency Of Lacto-*N*-Biose-I On The Growth Of Bifidobacteria As Compared To *A. oris*

The effect of lacto-N-biose I on *A. oris* growth was observed in vitro in the current study as was previously prebiotic effect was investigated in detail on bifidobacteria (Kiyohara *et al.*, 2009). The growth was tested on semisynthetic media containing lacto-N-biose I as a sole carbon source in the presence (0.5%) or absence of LNB in the media. Glucose was used as a positive control in place of LNB and the OD was taken at different time points in triplicate. Bifidobacterial strains showed good growth rate in lactose-supplemented media as compared to glucose-supplemented media suggesting that the LNB acts as a preferred source of energy over glucose in the growth of bifidobacteria (Kiyohara *et al.*, 2009). *B. bifidum* JCM1254, JCM7004, *B. breve* JCM1192, *B. longum* subsp. *infantis* JCM1210, JCM1222, and *B. longum* subsp. *longum* JCM1217, JCM7054 grew in the presence of LNB in the media and the growth was larger in LNB-supplemented media as compared to glucose-supplemented media. In the current study the growth of *A. oris* in LNB-supplemented media is pronounced and turbid cultures were obtained after a few hours of growth but they grew more in glucose-supplemented media, showing that LNB is not a preferred source of carbon but if present these *A. oris* have the ability to utilize LNB in vitro suggesting the adaptation of these strains to the human oral environmental conditions or utilization of different energy sources efficiently. This sheds light on the ecological fitness of *A. oris* in the human oral environment. Another possible reason of utilizing LNB-Type I by *A. oris* is that the oral cavity contains mucin for the most of time which has LNB as a structural unit, therefore, *A. oris* when present in plaque would utilizes LNB as a main energy source because milk oligosaccharides cannot remain in the mouth for a long time and flushes from the mouth either into the intestine or out of the mouth unless we keep cheese in the mouth for the most of time. The discussion suggest that the utilization of LNB-I was the preferred source of energy for *A. oris* as compared to glucose. *A. oris* strains grew and ferment better in presence of glucose but due to internal environmental conditions the main source of carbon is LNB from mucins, therefore the *A. oris* better utilizes LNB as

being adapted in the oral environment. In conclusion, LNB-type I is natural prebiotic and highly specific for the species.

The genome of *B. bifidum* PRL2010 was sequenced completely (Turroni *et al.*, 2010) and the LNB/GNB pathway was studied in detail. The deamination of *N*-acetylglucosamine and *N*-acetylgalactosamine was used which are present in mucin and other human- or diet-related hexosamines (HMOs). The genome of *B. bifidum* PRL2010 was also found to contain 20 ABC transporter proteins which are used to transport the oligosaccharides. In another study in 2007, a large selection of bifidobacterial strains were selected for their ability to utilize mucin as their sole energy source (Ventura *et al.*, 2007).

5.8.4 Genomics Of Host Glycan Utilization

In one of the studies, *B. bifidum* PRL2010 showed the best growth on mucin therefore this strain was selected to study mucin metabolism (Turroni *et al.*, 2010). Mucin acts as a semipermeable barrier and remains on the epithelial surfaces. Mucin has adhesive properties that is used by the gut microbiota to colonize (Johansson *et al.*, 2010). Mucin also contains a nutritional reservoir of carbohydrates. The main structural features contains O-linked and N-linked glycosylated domains in glycoproteins. N-glycans of oligosaccharides attached to *N*-acetylglucosamines (Gl*c*NA*c*) and O-linked oligosaccharides were found to attach with O-linked *N*-acetylgalactosamine (GalNAc) (Jensen *et al.*, 2010). Glycosyl hydrolases e.g endo-α-*N*-acetylgalactosaminidase (BBPR_02640) were encoded in the genomes which were used to degrade the oligosaccharides (Ashida *et al.*, 2008; Fujita *et al.*, 2005; Katayama *et al.*, 2005). An operon responsible for GNB and LNB utilization is present on the genome of *B. bifidum* PRL2010 (BBPR_1050-BBPR_1058). This operon consist of ABC-type oligosaccharide uptake system (BBPR_1056-BBPR_1058), and a gene encoding for LNBP used to cleave the oligosaccharides components LNB/GNB (Nishimoto & Kitaoka, 2007b). This is an important finding in which a LNB/GNB utilization pathway was found in *B. bifidum* which is different from *B. longum* subsp. *infantis* ATCC15697 (Sela *et al.*,

2008). *B. bifidum* used to uptake LNB/GNB from mucin and HMO with the help of extracellular enzymes 1,2-α-$_{L}$-fucosidases (BBPR_0193) and α-1,3/4-fucosidase (BBPR_1360). The role of these enzymes is to accelerate the defucosylation of LNB/GNB-containing glycans. This in turn facilitates the further breakdown of LNB/GNB by lacto-N-biosidases (BBPR_1438), endo-α-*N*-acetylgalactosaminidase (BBPR_0264) (Rose & Voynow, 2006; Wada *et al.*, 2008). The presence of two converging metabolic pathways of mucin and HMOs utilization by this bacterium clarifies the findings (Wada *et al.*, 2008). Figure 5.27 showed the gene location diagram in bifidobacteria to highlight the significance of Lacto-*N*-biose metabolic pathway and its role.

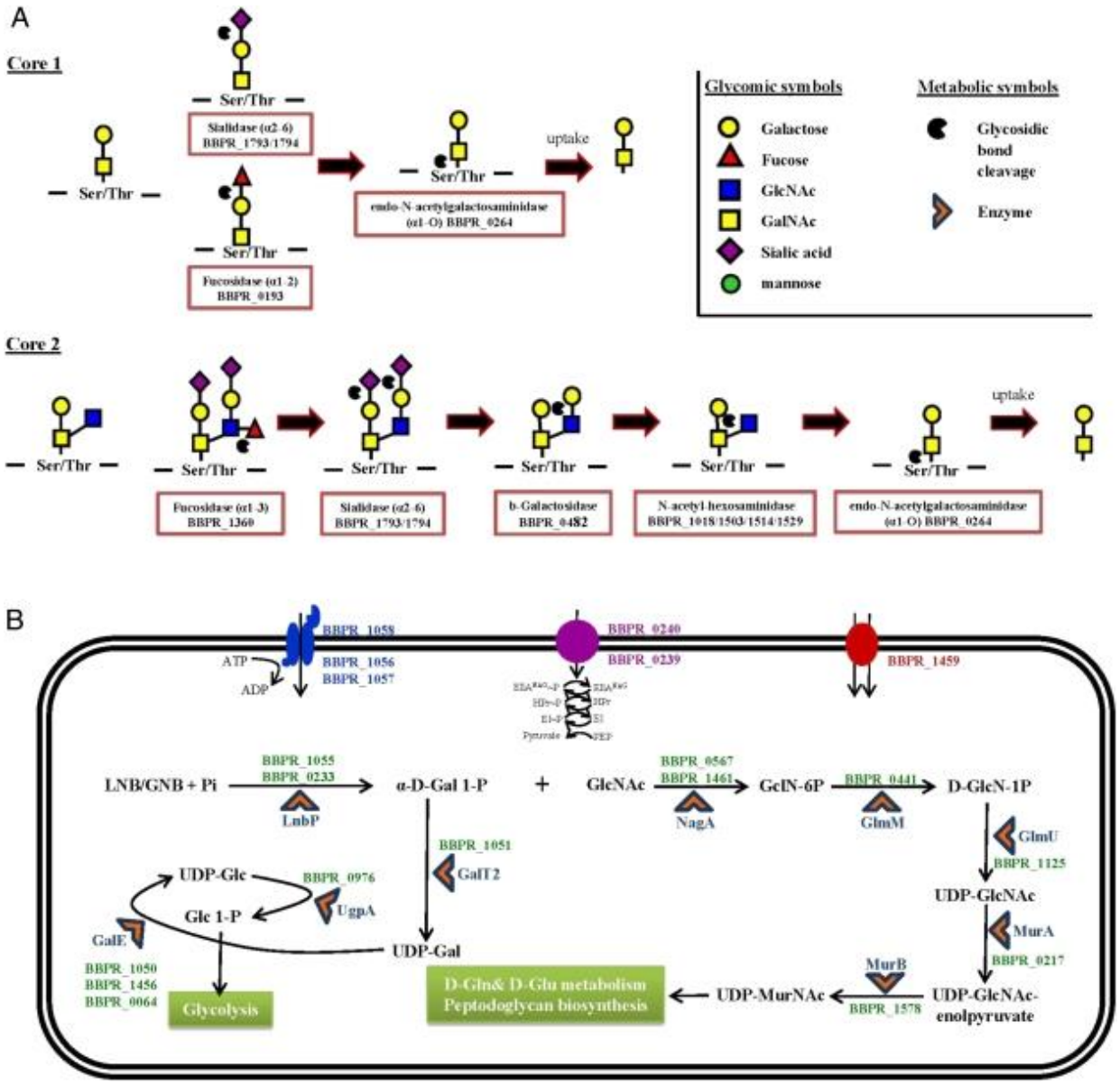


Figure 5.27: The gene locations in Bifidobacterium (Turroni et al., 2010).

***A represents target sites for the degradation of core 1 and core 2 O-glycans of mucins. Core 1 and core 2 are the major core structures of O-glycans of mucins.**

***B represents the metabolic pathway for catabolism of mucin oligosaccharides. Blue indicates transporter genes, violet indicates phosphotransferase systems and Red depicts symporter.**

5.8.5 RT-PCR Analysis Of *A. oris* And Adaptation To Mucin Utilization Compared To Proteome and Transcriptome Analysis of *B. bifidum* PRL2010

To prove that *A. oris* contains genes specific for utilization of mucins, RT-PCR analysis was carried out in which *A. oris* species were first grown in the presence of lactose-N-Biose I (building unit of mucin oligosaccharides) as their sole carbon source of the culture medium and then cDNA was synthesized from DNA and gene expression was observed as DNA bands using specifically designed primers and the experiments showed the up-regulation/ expression of the genes specific for LNB utilization in *A. oris* isolates. The proteome and transcriptome analysis of *B. bifidum* PRL2010 also supported the idea that when cells were grown on mucin greater expression of these genes was seen as compared to when grown only in the presence of lactose. This may be due to the presence of various substrates in mucins and HMOs which become available when grown on mucin and act as inducers of the LNB pathway-associated genes. Also putative aminopeptidases showed increased expression levels which seemed to be involved in the degradation of the proteinaceous component of mucin. 56 genes of *B. bifidum* PRL2010 showed increased expression when grown in the presence of mucin as their sole carbon source and some of these genes were also identified during proteome analysis (Turroni *et al.*, 2010).

In another study α-$_L$-fucosidase was prepared from *Streptomyces* spp. 142. This preparation was found to contain an enzyme activity against type I oligosaccharides releasing Lacto-*N*-biose from oligosacchardies thus showing the presence of Lacto-*N*-biosidase is essential to degrade Lacto-*N*-biose type I from oligosaccharides (Sano *et al.*, 1992).

5.8.6 Mutant Generation In Genes Of *A. oris*-MG1

In the current study, the mutant of *A. oris*-MG1 lacking the Ana_1637 gene was generated using in-frame deletion methodology using galK as a counterselection marker. A study by Cheng *et al*., (2012) developed an *A. oris* mutant using a markerless gene deletion system. mCherry fluorescence and resistance to kanamycin system was used to select for the mutants. Mishra *et al*., (2010) also developed an in- frame deletion system in *A. oris*-MG1 using galK as a counter selection marker and successfully employed the technique to generate mutants in *A. oris*-MG1.

5.9 Conclusion

A. oris and *A. naeslundii* belong to two distinct phylogenetic clusters. This is the first demonstration of Lacto-*N*-Biose utilization by the member of the oral microbiome. This may explain the greater prevalence of *A. oris* as compared to *A. naeslundii* in oral biofilms. *A. oris* has the ability to utilize Lacto-*N*-Biose while *A. naeslundii* do not.

Chapter 6 GENERAL DISCUSSION AND CONCLUSION

This work is divided into two parts of which the first part dealt with the genome sequencing of selected isolates of *Actinomyces oris* and *Actinomyces naeslundii*. The investigation of genomic diversity is very important with regard to finding the evolutionary history of bacterial species. The designation of isolates with correct position in phylogenetic tree has been of interest in oral microbiology since the early 20th century. *A. naeslundii* and *A. oris* groups have an important role in the development of dental plaque formation. However, the difficulties have been reported regarding clear identification of strains within *A. naeslundii* / *A. oris* group as they are highly related and could not be easily differentiated using 16S rDNA sequences. The oral micro-flora is more diverse than apparent from 16S sequencing studies. Therefore the lack of species identification was a limiting factor in studies which were interested in investigating the behaviour of these species. This limiting factor was addressed in the present study using recently developed high-throughput whole genome sequencing techniques. The complete genome sequencing of a large number of strains was not practical in the past but due to the advent of high-throughput sequencing technology, it is now possible to obtain the whole genome sequences. Therefore, in the first part, the main focus was on obtaining whole genome sequencing and investigation of the genetic diversity using bioinformatic approaches for the comparative genome analysis. Subsequently, the population dynamics of both groups was investigated using ClonalFrame and gene-by-gene sequence analysis. There is no doubt that the era of genome sequencing has had a vast and brisk effect on all fields of microbiology. Gene-by-gene sequencing studies were very difficult in the past for microbiologists but now with the advent of genome sequencing techniques, there is a relief in the scientific community from adapting the laborious and time-consuming tasks. The genomics also helps to understand the evolutionary history of bacteria, accelerates the microbial pathogenesis research work, and also opens areas of new research in the field of protein chemistry with the advent of a large number of unknown proteins and their unique folding patterns.

The current study has not only helped to expand the number of available genomes but also enabled the comparative genomic studies to link the genotype with the phenotype and also to investigate the biochemical characterisation of strains. Comparative

genomics will provide the baseline to implement the easily available molecular biological techniques in the laboratories. The current study used PCR-based methodology to find the differences in lacto-N-biose genes for larger number of strains. This will facilitate the discovery of particular gene without obtaining the complete genome sequences for larger number of strains. These approaches are simple, cheap and provide results in quick time and represent a short method to target specifically the gene of interest without obtaining the complete genomic sequences. For large scale comparison the same primer can be used to large numbers of test genomes.

The third Chapter describes the strategy of combining the short reads and long reads for the denovo assembly of high-throughput combined sequences. The quality of the denovo sequenced genomes was presented in detail and showed that de-novo assembly of high-throughput short and long read sequences was a very powerful method to get the high quality genome sequences of 36 *Actinomyces* strains.

The Genomics study described in chapter four obtained comparative analysis of *Actinomyces* strains, which included MLST analysis, Digital DNA-DNA hybridization studies, ClonalFrame analysis for detection of recombination and mutation events, and Core and Pan genome analysis of *A. oris* and *A. naeslundii* isolates. MLST analysis was carried out in which concatenation of seven house-keeping genes of publically available genomes of MG1, k20, c505, OT171, OT170, OT175 was added in the Neighbour-Joining tree of a study by Henssge (2009) to further investigate the true phylogenetic position of these strains. MG1, k20, c505 and OT175 fell into clusters of the main *A. oris* group while OT171 and OT170 were in separate phylogenetic positions. The DNA-DNA hybridization studies has also been used previously (Johnson's 1990) to investigate the phylogenetic status of *A. naeslundii*/viscous group but the method used was very difficult to implement due to variability of the results obtained in different laboratories. The current study used a digital DNA-DNA hybridization method, the results of which are based on availability of whole genome sequences and were authentic and reliable. Based on digital DNA-DNA hybridization, the isolates of *A. oris* group was divided into

six sub-groups and isolates of *A. naeslundii* were divided into three sub-groups. ClonalFrame analysis also sheds light on the population structure of *A. oris* and *A. naeslundii* group and the results are consistent with the results obtained with MLST analysis. There is also a novel method of obtaining core and pan genome, which was implemented in the current study. The core and pan genomes were obtained for 36 Actinomyces strains. 476 core genes were observed in 43 Actinomyces genomes. The Core genome size is smaller as compared to pan genome of these isolates showing the genomic diversity and adaptability to different environments.

Chapter five describes the characterisation of Lacto-*N*-biose operon and LNB utilization by *A. oris* isolates. Based on whole genome sequence data, *A. oris* strains appeared to possess a lacto-N-biose operon equivalent to that found in certain bifidobacteria. However, *A. naeslundii* strains appeared to have lost the operon. The function is more important than identity. In the larger second part of the work, the primers were designed to amplify the genes of interest (Lacto-*N*-biose) and all *A. oris* strains (n=25) harboured these genes while no *A. naeslundii* (n=25) did. In conclusion, the LNB operon was lost either in steps, sequentially or the LNB operon was lost at different times (more than once) and phylogenetic relationships may be established by whole genome comparisons therefore the species differences in the distribution of the LNB gene cluster were confirmed. There was no significant difference found among the growth values obtained using two different media while ASS medium proved to be a better medium because the *A. oris* growth was more profound and clear. *A. oris* and *A. naeslundii* belong to two distinct phylogenetic clusters. This is the first demonstration of Lacto-*N*-Biose utilization by a member of the oral microbiome. This surprising finding may suggest the greater prevalence of *A. oris* as compared to *A. naeslundii* in oral biofilms. *A. oris* has the ability to utilize Lacto-*N*-Biose while *A. naeslundii* does not. *A. oris* strains grew better on LNB than *A. naeslundii*.

The LNB operon region was compared in *A. naeslundii* with that of *A. oris* early in chapter 5. The next step was to prepare mutants of *A. oris*- MG1 in which the lacto-N-Biose phosphorylase gene was knocked out to investigate that whether this gene is

solely responsible for utilization of LNB or not. The third step included was to compare the growth of the parent strain to that of the mutant strain on media containing LNB to observe the growth pattern of mutant versus parent strain of *A. oris*-MG1 (Figure 6.1) .

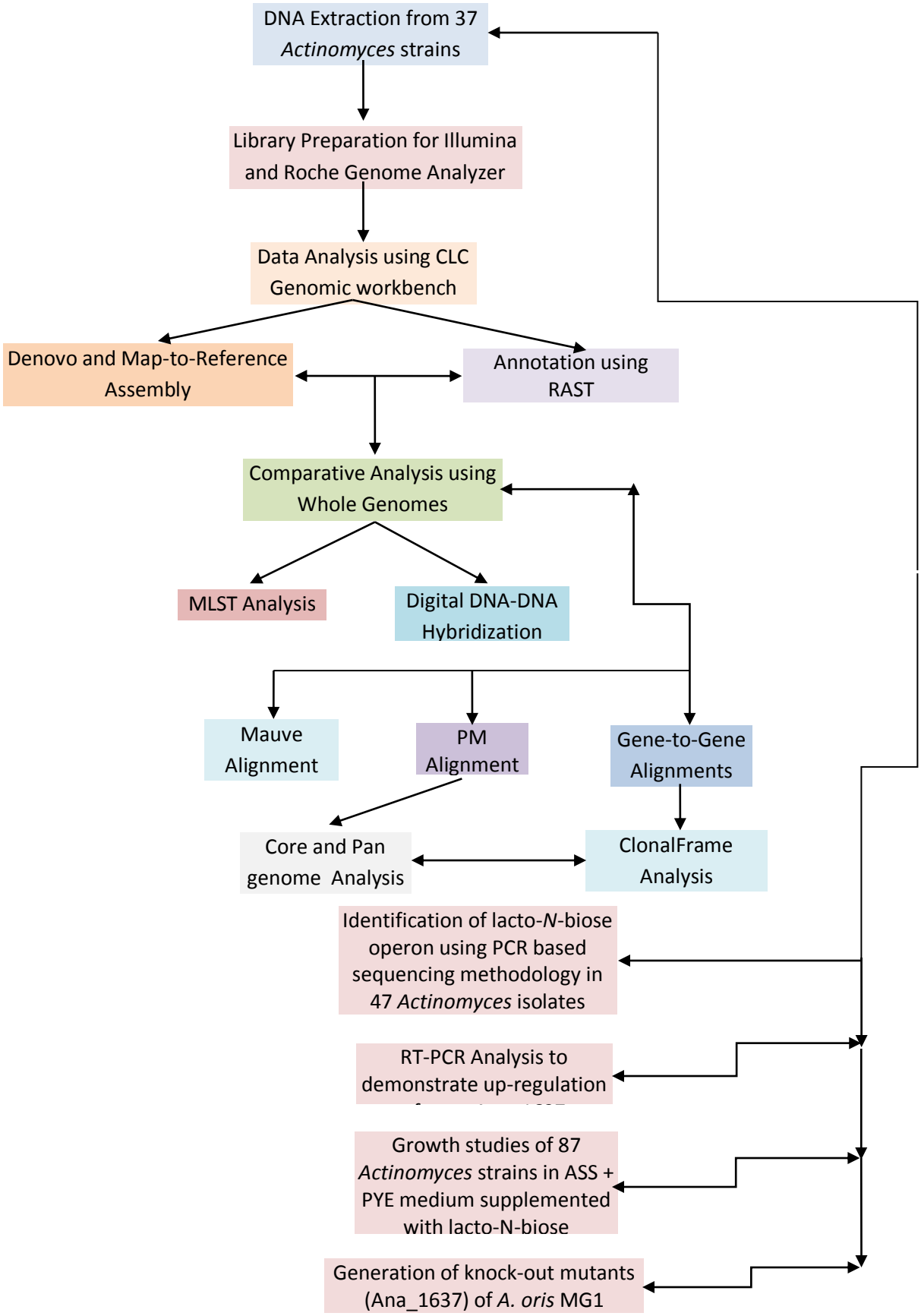


Figure 6.1: The work-flow diagram of the Material and methods used in current study. (PM: Progressive mauve; RT-PCR: Reverse transcriptase-PCR)

6.1 Next-Generation Sequencing

During the last five years, the incorporation of modern state-of-the art next generation sequencing techniques, which focused on the study of closely related genomes, helped to develop de-novo, high quality and finished assembly of a genome of interest (Brown, 2008). The emergence of this technique is cost effective and makes the high-throughput genome sequencing a reality (Marguerat, 2008). This instrument is capable of generating a million base pair reads from a single run. Currently two instruments are mostly used in microbial genetic studies and these are the Roche (454) GS FLX sequencer (http://www.454.com/enablingtechnology/the-system.asp) and the Illumina Genome Analyzer (http://www.illumina.com/pages.ilmn?ID=203). The work flow of the instruments is quite similar in principle except for a few changes (Shendure, 2008). The first step is to generate single stranded DNA. Libraries were constructed by fragmenting the genomic DNA. Adaptor sequences were ligated using simple laboratory based methods. The second main part is to generate clusters of DNA copies. For this purpose Roche 454 uses emulsion PCR. In this method library fragments were added into DNA capture beads. Each bead is capable of capturing a single molecule. The PCR reaction was carried out in water-in-oil micro-reactors. This results in one million copies of a single DNA fragment on each bead (Mardis, 2008). The other system is called Illumina Genome Analyzer, which uses bridge amplification methodology. In this system a flow cell is used which is a 8-channel solid device. The bridge amplification is carried out on the surface of the cell. Each DNA fragment resulted in clusters of DNA sequences which consist of more than 1,000 copies of the DNA molecule (Shendure, 2008).

After the generation of clonal clusters, the next step is known as sequencing-by-synthesis. Polymerase is required to initiate this step. The roche 454 system uses the pyrosequencing method. In this method the beads carrying amplicons were pre-incubated with *Bacillus stearothermophilus* (*bst*) DNA polymerase. In addition, other enzymes which were required are ATP sulfurylase, luciferase and apyrase, and the substrates adenosine 5´ phosphosulfate (APS) and luciferin. In this step, on each

nucleotide formation pyrophosphate (ppi) is released. The released PPi will be converted into ATP. Subsequently the substrates were added for the light production which were detected by sensors of the 454 system. The Illumina Genome analyzer uses the fluorescence scanning of each nucleotide. The read length produced by both systems is very variable. Roche 454 is capable to produce more than 400bp while Illumina can generate 76bp or 100 bp at the most. There is a limitation in the production of read length but still this system is better than the old technologies of sanger sequencing. The sequencing cost is also much reduced therefore the next generation technologies have much more potential for the production of the required results in cost effectiveness and time consumption (Mardis, 2008).

The impact of these techniques especially in the field of bacterial genomics has been widely accepted within the scientific community. Sequencing a whole genome was almost impossible in the past but now it is becoming a reality. There will be a time when the whole genome sequences of human, animals and micro-organisms will be publically available. The control of diseases would not remain a dream for the scientists. The 99% knowledge of bacterial diversity is still unknown today (Snyder, 2009) which will be revealed using the methods described in this thesis and also expanding the knowledge using many more new techniques in the near future. Furthermore the field of "metagenomics" has already initiated where the sequences of entire bacterial communities will be present. These bacterial communities may belong to different ecosystems therefore this project will help to explore the impact of environmental changes over time on the composition of the microbial species within the community (Snyder, 2009)

6.2 Significance For Clinical Microbiology

This work was another step forward in the improvement of taxonomy of human strains of *A. naeslundii*/viscosus. This work now has led to the description of *A. oris*, *A. naeslundii* and *A. johnsonii* as a distinct species. This finding will enable the clinical microbiologist as well as scientists to diagnose the clinical strains with authentic results, in less time and in a cost effective way. The methodology of growing strains in a medium containing oligosaccharides may be applied in clinical microbiology laboratories to define *A. oris* and *A. naeslundii*. Many laboratories already used 16S rRNA gene sequencing techniques and they may use LNBP primers described in this study for species identification.

6.3 Main Significance of Findings

I. *A. naeslundii* is a single group
II. *A. oris* appears to consist of second genetically distinct group, which may be either sub-species or distinct species.
III. This is supported by higher mutation rate in *A. oris***.**
IV. Core genome of *A. oris* is larger than *A. naeslundii*
V. *A. oris* has lacto-*N*-biose metabolic pathway, which may explain its numerical predominance over *A. naeslundii* in the mouth.

6.4 Final Conclusion

I. The nearly complete whole genome sequencing of 36 Actinomyces strains clarified the ambiguous taxonomy of human *A. naeslundii*/viscosus group and *A. oris*, *A. naeslundii* and *A. johnsonii* were kept as it is as were classified previously using MLST analysis.

II. The De-novo assembly of High-throughput short-reads and long reads sequences were obtained

III. The members of the main *A. oris* group cannot be grouped as the same species as revealed by Digital DNA-DNA hybridization analysis study.

IV. The operon for Lacto-N-biose utilization presence in *A. oris* and absence in *A. naeslundii* was for the first time reported in the current study.

V. The *A. oris* and *A. naeslundii* have a clonal population structure and recombination/mutation events are greater in *A. oris* as compared to *A. naeslundii* isolates.

6.5 Future Work

This work can be taken further by doing the RNA-sequencing of the selected strains of Actinomyces to investigate which genes are over-expressed in one group of Actinomyces as compared to another group of strains. RNA-sequencing is a very powerful tool for the in-depth analysis of the genes and their role in controlling certain features especially colonization strategy for forming plaque and biofilms. Biological meaning of the sequence differences in certain genes can be very helpful to elucidate further the role behind the assembly and colonization on the teeth and other body parts. Based on the findings in the current study that *A. oris* has the ability to utilize lacto-*N*-biose while *A. naeslundii* do not, the next step should be to investigate the metabolomics of the enzymes involved in this metabolic pathway. To determine the metabolic intermediates of LNB metabolism by *A. oris* and *A. naeslundii* with galactose and N-acetylglucosamine as controls would be an important finding to exactly determine the role of intermediate genes/enzymes and the quantity they produced of LNB. As a result of metabolic process it would enable us to determine the extent of utilization of LNB as a primary source of sugars for the growth and proliferation of *A. oris* in dental plaque as compared to *A. naeslundii*. Binding assays to several salivary proteins and their properties such as binding to other isolates of plaque forming bacteria would be of interest especially in the context of the utilization of LNB. Another aspect would be to determine the relationship of important genes and proteins to other actinomyces and bacteria to examine the mechanisms involved for the colonization in similar habitats. Genetic-engineering of the important genes to understand the mechanisms and role of the particular genes would be of particular interest as we prepared mutant of *A. oris*-lacking the LNB gene and then investigated further that mutant was unable to grow in the presence of LNB. This strategy can also be used to unravel the mechanisms of other metabolically important genes and their role.

Further research is clearly needed to understand the mechanism and biological meaning of the genes involved and reveal its importance in a clinical environments. This work has provided the basis for clinical microbiologists and molecular biologists to achieve

these targets as nearly complete genomic sequences would be available for the scientific community publically.

7. REFERENCES

Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I. & Dewhirst, F. E. (2005). Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* **43**, 5721-5732.

Achtman, M. & Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. *Nat Rev Micro* **6**, 431-440.

Alam, S., Brailsford, S. R., Whiley, R. A. & Beighton, D. (1999). PCR-Based methods for genotyping viridans group streptococci. *J Clin Microbiol* **37**, 2772-2776.

Alam, S., Brailsford, S. R., Adams, S., Allison, C., Sheehy, E., Zoitopoulos, L., Kidd, E. A. & Beighton, D. (2000). Genotypic heterogeneity of Streptococcus oralis and distinct aciduric subpopulations in human dental plaque. *Appl Environ Microbiol* **66**, 3330-3336.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Andersen, R. N., Ganeshkumar, N. & Kolenbrander, P. E. (1998). Helicobacter pylori adheres selectively to Fusobacterium spp. *Oral Microbiol Immunol* **13**, 51-54.

Anderson, T. J. C., Haubold, B., Williams, J. T. & other authors (2000). Microsatellite Markers Reveal a Spectrum of Population Structures in the Malaria Parasite Plasmodium falciparum. *Molecular Biology and Evolution* **17**, 1467-1482.

Aparna, M. S. & Yadav, S. (2008). Biofilms: microbes and disease. *Brazilian Journal of Infectious Diseases* **12**, 526-530.

Arif, N., Sheehy, E. C., Do, T. & Beighton, D. (2008). Diversity of Veillonella spp. from sound and carious sites in children. *J Dent Res* **87**, 278-282.

Ashida, H., Maki, R., Ozawa, H. & other authors (2008). Characterization of two different endo-alpha-N-acetylgalactosaminidases from probiotic and pathogenic enterobacteria, Bifidobacterium longum and Clostridium perfringens. *Glycobiology* **18**, 727-734. doi: 710.1093/glycob/cwn1053. Epub 2008 Jun 1017.

Auch, A. F., Klenk, H.-P. & GÃ¶ker, M. (2010a). *Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs*.

Auch, A. F., von Jan, M., Klenk, H.-P. & GÃ¶ker, M. (2010b). *Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison*.

Barsotti, O., Morrier, J. J., Decoret, D., Benay, G. & Rocca, J. P. (1993). An investigation into the use of restriction endonuclease analysis for the study of transmission of Actinomyces. *J Clin Periodontol* **20**, 436-442.

Beatty, P., Hanisch, F. G., Stolz, D. B., Finn, O. J. & Ciborowski, P. (2001). Biochemical characterization of the soluble form of tumor antigen MUC1 isolated from sera and ascites fluid of breast and pancreatic cancer patients. *Clin Cancer Res* **7**, 781s-787s.

Beighton, D. & Smith, K. (1986). The modulation of exoglycosidic enzymes in the supragingival plaque of macaque monkeys. *FEMS Microbiology Letters* **34**, 319-322.

Beighton, D., Smith, K. & Hayday, H. (1986). The growth of bacteria and the production of exoglycosidic enzymes in the dental plaque of macaque monkeys. *Arch Oral Biol* **31**, 829-835.

Beighton, D. & Whiley, R. A. (1990). Sialidase activity of the "Streptococcus milleri group" and other viridans group streptococci. *Journal of Clinical Microbiology* **28**, 1431-1433.

Beighton, D., Hardie, J. M. & Whiley, R. A. (1991a). A scheme for the identification of viridans streptococci. *Journal of Medical Microbiology* **35**, 367-372.

Beighton, D., Russell, R. R. & Whiley, R. A. (1991b). A simple biochemical scheme for the differentiation of Streptococcus mutans and Streptococcus sobrinus. *Caries Res* **25**, 174-178.

Beighton, D. (2005). The complex oral microflora of high-risk individuals and groups and its role in the caries process. *Community Dentistry and Oral Epidemiology* **33**, 248-255.

Beighton, D., Gilbert, S. C., Clark, D. & other authors (2008). Isolation and identification of bifidobacteriaceae from human saliva. *Appl Environ Microbiol* **74**, 6457-6460. doi: 6410.1128/AEM.00895-00808. Epub 02008 Aug 00822.

Beighton, D., Sadaf Rasheed Mughal and Thuy Do (2013). Bacterial Catabolism of Salivary Substrates. *Oral Microbial Ecology: Current research and New perspective*, 37-46.

Bennett, J. S., Jolley, K. A., Sparling, P. F., Saunders, N. J., Hart, C. A., Feavers, I. M. & Maiden, M. C. (2007). Species status of Neisseria gonorrhoeae: evolutionary and epidemiological inferences from multilocus sequence typing. *BMC Biol* **5**, 35.

Bezirtzoglou, E., Maipa, V., Chotoura, N., Apazidou, E., Tsiotsias, A., Voidarou, C., Kostakis, D. & Alexopoulos, A. (2006). Occurrence of Bifidobacterium in the intestine of newborns by fluorescence in situ hybridization. *Comp Immunol Microbiol Infect Dis* **29**, 345-352.

Bezkorovainy, A. (1989). Ecology of Bifidobactria *In "Biochemistry and Physiology of Bifidobacteria," eds Bezkorovainny, A, and Miller-Catchpole, R, CRC Press, Cleveland*, 29-72.

Bik, E. M., Long, C. D., Armitage, G. C. & other authors (2010). Bacterial diversity in the oral cavity of 10 healthy individuals. *Isme J* **4**, 962-974.

Binnewies, T. T., Motro, Y., Hallin, P. F. & other authors (2006). Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. *Funct Integr Genomics* **6**, 165-185.

Birtles, A., Hardy, K., Gray, S. J., Handford, S., Kaczmarski, E. B., Edwards-Jones, V. & Fox, A. J. (2005). Multilocus sequence typing of Neisseria meningitidis directly from clinical samples and application of the method to the investigation of meningococcal disease case clusters. *J Clin Microbiol* **43**, 6007-6014.

Bisharat, N., Cohen, D. I., Maiden, M. C., Crook, D. W., Peto, T. & Harding, R. M. (2007). The evolution of genetic structure in the marine pathogen, Vibrio vulnificus. *Infect Genet Evol* **7**, 685-693.

Bollinger, O. (1877). Über eine neue Pilzkrankheit beim
Rinde. Zbl. . . *Med Wissensch*, 481-485.

Bonnet, R. g., Suau, A., DorÃ©, J. l., Gibson, G. R. & Collins, M. D. (2002). Differences in rDNA libraries of faecal bacteria derived from 10- and 25-cycle PCRs. *International Journal of Systematic and Evolutionary Microbiology* **52**, 757-763.

Bos, R., van der Mei, H. C. & Busscher, H. J. (1996). Co-adhesion of oral microbial pairs under flow in the presence of saliva and lactose. *J Dent Res* **75**, 809-815.

Bouffard, G. G., Rudd, K. E. & Adhya, S. L. (1994). Dependence of lactose metabolism upon mutarotase encoded in the gal operon in Escherichia coli. *J Mol Biol* **244**, 269-278.

Bradshaw, D. J., Homer, K. A., Marsh, P. D. & Beighton, D. (1994). Metabolic cooperation in oral microbial communities during growth on mucin. *Microbiology* **140**, 3407-3412.

Bradshaw, D. J., Marsh, P. D., Watson, G. K. & Allison, C. (1998). Role of Fusobacterium nucleatum and coaggregation in anaerobe survival in planktonic and biofilm oral microbial communities during aeration. *Infect Immun* **66**, 4729-4732.

Braham, P. H. & Moncla, B. J. (1992). Rapid presumptive identification and further characterization of Bacteroides forsythus. *Journal of Clinical Microbiology* **30**, 649-654.

Brailsford, S. R., Lynch, E. & Beighton, D. (1998). The isolation of Actinomyces naeslundii from sound root surfaces and root carious lesions. *Caries Res* **32**, 100-106.

Brailsford, S. R., Tregaskis, R. B., Leftwich, H. S. & Beighton, D. (1999). The predominant Actinomyces spp. isolated from infected dentin of active root caries lesions. *J Dent Res* **78**, 1525-1534.

Brailsford, S. R., Shah, B., Simons, D., Gilbert, S., Clark, D., Ines, I., Adams, S. E., Allison, C. & Beighton, D. (2001). The predominant aciduric microflora of root-caries lesions. *J Dent Res* **80**, 1828-1833.

Bratt, P., Boren, Boren, T. & Str omberg, N. (1999). Secretory immunoglobulin A heavy chain presents Galbeta1-3GalNAc binding structures for Actinomyces naeslundii genospecies 1. *J Dent Res* **78**, 1238-1244.

Brenner, S., Johnson, M., Bridgham, J. & other authors (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotech* **18**, 630-634.

Brown, J. R. (2008). Comparative Genomics: Basic and Applied Research. . *Boca Raton: CRC Press, Taylor & Francis Group*.

Byers, H. L., Tarelli, E., Homer, K. A. & Beighton, D. (2000). Isolation and characterisation of sialidase from a strain of Streptococcus oralis. *J Med Microbiol* **49**, 235-244.

Chaves, D., Sandoval, A., Rodriguez, L., Garcia, J. C., Restrepo, S. & Zambrano, M. M. (2010). [Comparative analysis of six Mycobacterium tuberculosis complex genomes]. *Biomedica* **30**, 23-31.

Cheaib, Z. & Lussi, A. (2013). Role of amylase, mucin, IgA and albumin on salivary protein buffering capacity: A pilot study. *J* **38**, 259-265.

Chen, C., Zhang, W., Zheng, H. & other authors (2013). Minimum Core Genome Sequence Typing of Bacterial Pathogens: a Unified Approach for Clinical and Public Health Microbiology. *Journal of Clinical Microbiology* **51**, 2582-2591.

Chen, P., Cisar, J. O., Hess, S., Ho, J. T. & Leung, K. P. (2007). Amended description of the genes for synthesis of Actinomyces naeslundii T14V type 1 fimbriae and associated adhesin. *Infect Immun* **75**, 4181-4185.

Chen, T., Yu, W.-H., Izard, J., Baranova, O. V., Lakshmanan, A. & Dewhirst, F. E. (2010). The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database* **2010**.

Cheon, K., Moser, S. A., Whiddon, J., Osgood, R. C., Momeni, S., Ruby, J. D., Cutter, G. R., Allison, D. B. & Childers, N. K. (2011). Genetic diversity of plaque mutans streptococci with rep-PCR. *J Dent Res* **90**, 331-335.

Cho, J.-C. & Tiedje, J. M. (2001). Bacterial Species Determination from DNA-DNA Hybridization by Using Genome Fragments and DNA Microarrays. *Applied and Environmental Microbiology* **67**, 3677-3682.

Choi, B. K., Paster, B. J., Dewhirst, F. E. & GÃ¶bel, U. B. (1994). Diversity of cultivable and uncultivable oral spirochetes from a patient with severe destructive periodontitis. *Infection and Immunity* **62**, 1889-1895.

Choi, B. K., Wyss, C. & GÃ¶bel, U. B. (1996). Phylogenetic analysis of pathogen-related oral spirochetes. *Journal of Clinical Microbiology* **34**, 1922-1925.

Christensen, H., Kuhnert, P., Olsen, J. E. & Bisgaard, M. (2004). Comparative phylogenies of the housekeeping genes atpD, infB and rpoB and the 16S rRNA gene within the Pasteurellaceae. *Int J Syst Evol Microbiol* **54**, 1601-1609.

Cisar, J. O., Kolenbrander, P. E. & McIntire, F. C. (1979). Specificity of coaggregation reactions between human oral streptococci and strains of Actinomyces viscosus or Actinomyces naeslundii. *Infection and Immunity* **24**, 742-752.

Cisar, J. O., Sandberg, A. L., Abeygunawardana, C., Reddy, G. P. & Bush, C. A. (1995). Lectin recognition of host-like saccharide motifs in streptococcal cell wall polysaccharides. *Glycobiology* **5**, 655-662.

Cisar, J. O., Takahashi, Y., Ruhl, S., Donkersloot, J. A. & Sandberg, A. L. (1997). Specific inhibitors of bacterial adhesion: observations from the study of gram-positive bacteria that initiate biofilm formation on the tooth surface. *Adv Dent Res* **11**, 168-175.

Clarridge, J. E. (2004). Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. *Clinical Microbiology Reviews* **17**, 840-862.

Coenye, T., Gevers, D., Van de Peer, Y., Vandamme, P. & Swings, J. (2005). Towards a prokaryotic genomic taxonomy. *FEMS Microbiol Rev* **29**, 147-167.

Cohan, F. M. (2002). What are bacterial species? *Annu Rev Microbiol* **56**, 457-487. Epub 2002 Jan 2030.

Cole, J. R., Wang, Q., Cardenas, E. & other authors (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**, D141-145. doi: 110.1093/nar/gkn1879. Epub 2008 Nov 1012.

Colwell, R. R., R.A. Clayton, B.A. Ortiz-Condel, D. Jacobs, and & Russek-Cohen., E. (1995). The microbial species concept and biodiversity. In D. Allsopp, R.R. Colwell, and D.L. Hawksworth (eds.)
Microbial Diversity and Ecosystem Function
Oxford: CAB International, 3-15.

Coppa, G. V., Bruni, S., Morelli, L., Soldi, S. & Gabrielli, O. (2004). The first prebiotics in humans: human milk oligosaccharides. *J Clin Gastroenterol* **38**, S80-83.

Costello, A. H., Cisar, J. O., Kolenbrander, P. E. & Gabriel, O. (1979). Neuraminidase-dependent hamagglutination of human erythrocytes by human strains of Actinomyces viscosus and Actinomyces naeslundii. *Infection and Immunity* **26**, 563-572.

Costerton, J. W., Geesey, G. G. & Cheng, K. J. (1978). How bacteria stick. *Sci Am* **238**, 86-95.

Coykendall, A. L. & Munzenmaier, A. J. (1979). Deoxyribonucleic Acid Hybridization Among Strains of Actinomyces viscosus and Actinomyces naeslundii. *Int J Syst Bacteriol* **29**, 234-240.

Curtis, T. P., Sloan, W. T. & Scannell, J. W. (2002). Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences* **99**, 10494-10499.

Darling, A. E., Miklos, I. & Ragan, M. A. (2008). Dynamics of genome rearrangement in bacterial populations. *PLoS Genet* **4**, e1000128. doi: 1000110.1001371/journal.pgen.1000128.

Darling, A. E., Mau, B. & Perna, N. T. (2010). progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* **5**, e11147.

Davey, M. E. & Costerton, J. W. (2006). Molecular genetics analyses of biofilm formation in oral isolates. *Periodontology 2000* **42**, 13-26.

de Lillo, A., Ashley, F. P., Palmer, R. M., Munson, M. A., Kyriacou, L., Weightman, A. J. & Wade, W. G. (2006). Novel subgingival bacterial phylotypes detected using multiple universal polymerase chain reaction primer sets. *Oral Microbiology and Immunology* **21**, 61-68.

Deloger, M., El Karoui, M. & Petit, M. A. (2009). A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. *J Bacteriol* **191**, 91-99. doi: 10.1128/JB.01202-01208. Epub 02008 Oct 01231.

Derensy-Dron, D., Krzewinski, F., Brassart, C. & Bouquelet, S. (1999). Beta-1,3-galactosyl-N-acetylhexosamine phosphorylase from Bifidobacterium bifidum DSM 20082: characterization, partial purification and relation to mucin degradation. **29 (Pt 1)**, 3-10.

Dethlefsen, L., McFall-Ngai, M. & Relman, D. A. (2007). An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* **449**, 811-818.

Dewhirst, F., Chen, T., Izard, J., Paster, B. J., Tanner, A. C. R., Yu, W.-h., Lakshmanan, A. & Wade, W. G. (2010). The Human Oral Microbiome. *Journal of Bacteriology*.

Diaz, P. I., Chalmers, N. I., Rickard, A. H., Kong, C., Milburn, C. L., Palmer, R. J. & Kolenbrander, P. E. (2006). Molecular Characterization of Subject-Specific Oral Microflora during Initial Colonization of Enamel. *Applied and Environmental Microbiology* **72**, 2837-2848.

Didelot, X. & Falush, D. (2007). Inference of Bacterial Microevolution Using Multilocus Sequence Data. *Genetics* **175**, 1251-1266.

Dige, I., Nyengaard, J. R., Kilian, M. & Nyvad, B. (2009). Application of stereological principles for quantification of bacteria in intact dental biofilms. *Oral Microbiology and Immunology* **24**, 69-75.

Dige, I., Sebastian Schlafer,Bente Nyvad (2012). Difference in initial dental biofilm accumulation between night and day. *Acta Odontologica Scandinavica* **70**, 441-447.

Dingle, K. E., Colles, F. M., Wareing, D. R. & other authors (2001). Multilocus sequence typing system for Campylobacter jejuni. *J Clin Microbiol* **39**, 14-23.

Do, T., Henssge, U., Gilbert, S. C., Clark, D. & Beighton, D. (2008). Evidence for recombination between a sialidase (nanH) of Actinomyces naeslundii and Actinomyces oris, previously named 'Actinomyces naeslundii genospecies 1 and 2'. *FEMS Microbiology Letters* **288**, 156-162.

Do, T., Gilbert, S. C., Clark, D. & other authors (2010). Generation of diversity in Streptococcus mutans genes demonstrated by MLST. *PLoS One* **5**, 0009073.

Donlan, R. M. & Costerton, J. W. (2002). Biofilms: Survival Mechanisms of Clinically Relevant Microorganisms. *Clinical Microbiology Reviews* **15**, 167-193.

Douglas, C. W. (1994). Bacterial-protein interactions in the oral cavity. *Adv Dent Res* **8**, 254-262.

Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E., Schaal, K., Yassin, A. & Stackebrandt, E. (2006). The Family Actinomycetaceae: The Genera Actinomyces, Actinobaculum, Arcanobacterium, Varibaculum, and Mobiluncus. In *The Prokaryotes*, pp. 430-537: Springer New York.

Dykhuizen, D. (2005). Species Numbers in Bacteria. *Proc Calif Acad Sci* **56**, 62-71.

Dymock, D., Weightman, A. J., Scully, C. & Wade, W. G. (1996). Molecular analysis of microflora associated with dentoalveolar abscesses. *J Clin Microbiol* **34**, 537-542.

Ellen, R. P. & Balcerzak-Raczkowski, I. B. (1975). Differential medium for detecting dental plaque bacteria resembling Actinomyces viscosus and Actinomyces naeslundii. *Journal of Clinical Microbiology* **2**, 305-310.

Ellen, R. P. (1976). Establishment and distribution of Actinomyces viscosus and Actinomyces naeslundii in the human oral cavity. *Infection and Immunity* **14**, 1119-1124.

Ellen, R. P., Fillery, E. D., Chan, K. H. & Grove, D. A. (1980). Sialidase-Enhanced Lectin-Like Mechanism for Actinomyces viscosus and Actinomyces naeslundii Hemagglutination. *Infection and Immunity* **27**, 335-343.

Enright, M. C. & Spratt, B. G. (1998). A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. *Microbiology* **144**, 3049-3060.

Enright, M. C., Day, N. P., Davies, C. E., Peacock, S. J. & Spratt, B. G. (2000). Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of Staphylococcus aureus. *J Clin Microbiol* **38**, 1008-1015.

Erikson, D. (1940). Pathogenic anaerobic organisms of the Actinomyces group.
. *ERIKSOND, (1940) Pathogenic anaerobic organisms of the Actinomyces group Spec Rep Ser med Res Coun, Lond* **240**, 5-63.

Evaldson, G., Heimdahl, A., Kager, L. & Nord, C. E. (1982). The normal human anaerobic microflora. *Scand J Infect Dis Suppl* **35**, 9-15.

Ezaki, T., Hashimoto, Y. & Yabuuchi, E. (1989). Fluorometric Deoxyribonucleic Acid-Deoxyribonucleic Acid Hybridization in Microdilution Wells as an Alternative to Membrane Filter Hybridization in which Radioisotopes Are Used To Determine Genetic Relatedness among Bacterial Strains. *Int J Syst Bacteriol* **39**, 224-229.

Feil, E. J., Smith, J. M., Enright, M. C. & Spratt, B. G. (2000). Estimating recombinational parameters in Streptococcus pneumoniae from multilocus sequence typing data. *Genetics* **154**, 1439-1450.

Feil, E. J. & Spratt, B. G. (2001). Recombination and the population structures of bacterial pathogens. *Annu Rev Microbiol* **55**, 561-590.

Fejerskov, O. & Kidd, E. A. M. (2008). *Dental caries : the disease and its clinical management*. Oxford; Ames, Iowa: Blackwell Munksgaard.

Felsenstein, J. (1985). Confidence-limits on Phylogeneis- an approach using the bootstrap. *Evolution* **39**, 783-791.

Fillery, E. D., Bowden, G. H. & Hardie, J. M. (1978). A comparison of strains of bacteria designated Actinomyces viscosus and Actinomyces naeslundii. *Caries Res* **12**, 299-312.

Fleischmann, R. D., Adams, M. D., White, O. & other authors (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269**, 496-512.

Flores, G. E., Henley, J. B. & Fierer, N. (2012). A direct PCR approach to accelerate analyses of human-associated microbial communities. *PLoS One* **7**, 4.

Frandsen, E. V. (1994). Carbohydrate depletion of immunoglobulin A1 by oral species of gram-positive rods. *Oral Microbiol Immunol* **9**, 352-358.

Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G. & Hanage, W. P. (2009). The bacterial species challenge: making sense of genetic and ecological diversity. *Science* **323**, 741-746.

Fraser, C. M., Eisen, J., Fleischmann, R. D., Ketchum, K. A. & Peterson, S. (2000). Comparative genomics and understanding of microbial biology. *Emerg Infect Dis* **6**, 505-512.

Frey, P. A. (1996). The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. *Faseb J* **10**, 461-470.

Fujita, K., Oura, F., Nagamine, N., Katayama, T., Hiratake, J., Sakata, K., Kumagai, H. & Yamamoto, K. (2005). Identification and molecular cloning of a novel glycoside hydrolase family of core 1 type O-glycan-specific endo-alpha-N-acetylgalactosaminidase from Bifidobacterium longum. *J Biol Chem* **280**, 37415-37422. Epub 32005 Sep 37411.

Garrido, D., Ruiz-Moyano, S. & Mills, D. A. (2012). Release and utilization of N-acetyl-D-glucosamine from human milk oligosaccharides by Bifidobacterium longum subsp. infantis. *Anaerobe* **18**, 430-435.

Georg, L. K., L Pine, and M. A. Gerenser. (1969). Actinomyces viscosus, comb. nov., a catalase-positive, facultative
member of the genus Actinomyces. *Int J Syst Bacteriol*
19, 291-293.

Gerencser, M. A. & Slack, J. M. (1976). Serological identification of Actinomyces using fluorescent antibody techniques. *J Dent Res* **55**, A184-191.

Gerencser, M. A. & Norris, T. B. a. J. R. (1979). Chapter VI The Application of Fluorescent Antibody Techniques to the Identification of Actinomyces and Arachnia. In *Methods in Microbiology*, pp. 287-321: Academic Press.

Gevers, D., Cohan, F. M., Lawrence, J. G. & other authors (2005). Re-evaluating prokaryotic species. *Nat Rev Micro* **3**, 733-739.

Gevers, D., Knight, R., Petrosino, J. F. & other authors (2012). The Human Microbiome Project: a community resource for the healthy human microbiome. *PLoS Biol* **10**, e1001377.

Gibbons, R. J. & Nygaard, M. (1970). Interbacterial aggregation of plaque bacteria. *Arch Oral Biol* **15**, 1397-1400.

Gibbons, R. J., Etherden, I. & Moreno, E. C. (1983). Association of neuraminidase-sensitive receptors and putative hydrophobic interactions with high-affinity binding sites for Streptococcus sanguis C5 in salivary pellicles. *Infection and Immunity* **42**, 1006-1012.

Gibbons, R. J., Hay, D. I., Cisar, J. O. & Clark, W. B. (1988). Adsorbed salivary proline-rich protein 1 and statherin: receptors for type 1 fimbriae of Actinomyces viscosus T14V-J1 on apatitic surfaces. *Infection and Immunity* **56**, 2990-2993.

Gibbons, R. J., Hay, D. I., Childs, W. C., 3rd & Davis, G. (1990). Role of cryptic receptors (cryptitopes) in bacterial adhesion to oral surfaces. *Arch Oral Biol* **35**, 107S-114S.

Giovannoni, S. J. & Stingl, U. (2005). Molecular diversity and ecology of microbial plankton. *Nature* **437**, 343-348.

Gogarten, J. P. & Townsend, J. P. (2005). Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* **3**, 679-687.

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P. & Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**, 81-91.

Grimont, P. A. D., Popoff, M. Y., Grimont, F., Coynault, C. & Lemelin, M. (1980). Reproducibility and correlation study of three deoxyribonucleic acid hybridization procedures. *Current Microbiol* **4**, 325-330.

Gusberti, F. A., Gada, T. G., Lang, N. P. & Geering, A. H. (1985). Cultivable microflora of plaque from full denture bases and adjacent palatal mucosa. *J Biol Buccale* **13**, 227-236.

Haarman, M. & Knol, J. (2005). Quantitative real-time PCR assays to identify and quantify fecal Bifidobacterium species in infants receiving a prebiotic infant formula. *Appl Environ Microbiol* **71**, 2318-2324.

Haffajee, A. D., Socransky, S. S., Patel, M. R. & Song, X. (2008). Microbial complexes in supragingival plaque. *Oral Microbiology and Immunology* **23**, 196-205.

Hall, V., O'Neill, G. L., Magee, J. T. & Duerden, B. I. (1999). Development of amplified 16S ribosomal DNA restriction analysis for identification of Actinomyces species and comparison with pyrolysis-mass spectrometry and conventional biochemical tests. *J Clin Microbiol* **37**, 2255-2261.

Hall, v., Brazier, j. s. (1997). Identification of actinomyces—what are the major problems? in Anaerobic pathogens. *eds Eley A R, Bennett K W (Academic Press, Sheffield, England)*, 187–192.

Hallberg, K., Holm, C., Ohman, U. & Stromberg, N. (1998). Actinomyces naeslundii displays variant fimP and fimA fimbrial subunit genes corresponding to different types of acidic proline-rich protein and beta-linked galactosamine binding specificity. *Infect Immun* **66**, 4403-4410.

Hamada, S. & Slade, H. D. (1980). Biology, immunology, and cariogenicity of Streptococcus mutans. *Microbiol Rev* **44**, 331-384.

Hanage, W. P., Fraser, C. & Spratt, B. G. (2005). Fuzzy species among recombinogenic bacteria. *BMC Biol* **3**, 6.

Hardie, J. M. (1992). Oral microbiology: current concepts in the microbiology of dental caries and periodontal disease. *Br Dent J* **172**, 271-278.

Harmsen, H. J., Wildeboer-Veloo, A. C., Raangs, G. C., Wagendorp, A. A., Klijn, N., Bindels, J. G. & Welling, G. W. (2000). Analysis of intestinal flora development in breast-fed and formula-fed infants by using molecular identification and detection methods. *J Pediatr Gastroenterol Nutr* **30**, 61-67.

Harper-Owen, R., Dymock, D., Booth, V., Weightman, A. J. & Wade, W. G. (1999). Detection of unculturable bacteria in periodontal health and disease by PCR. *J Clin Microbiol* **37**, 1469-1473.

He, X., Wu, C., Yarbrough, D., Sim, L., Niu, G., Merritt, J., Shi, W. & Qi, F. (2008). The cia operon of Streptococcus mutans encodes a unique component required for calcium-mediated autoregulation. *Mol Microbiol* **70**, 112-126. doi: 110.1111/j.1365-2958.2008.06390.x. Epub 02008 Aug 06394.

Henssge, U., Do, T., Radford, D. R., Gilbert, S. C., Clark, D. & Beighton, D. (2009). Emended description of Actinomyces naeslundii and descriptions of Actinomyces oris sp. nov. and Actinomyces johnsonii sp. nov., previously identified as Actinomyces

naeslundii genospecies 1, 2 and WVA 963. *International Journal of Systematic and Evolutionary Microbiology* **59**, 509-516.

Henssge, U., Do, T., Gilbert, S. C., Cox, S., Clark, D., Wickstrom, C., Ligtenberg, A. J., Radford, D. R. & Beighton, D. (2011). Application of MLST and pilus gene sequence comparisons to investigate the population structures of Actinomyces naeslundii and Actinomyces oris. *PLoS One* **6**, 30.

Henz, S. R., Huson, D. H., Auch, A. F., Nieselt-Struwe, K. & Schuster, S. C. (2005). Whole-genome prokaryotic phylogeny. *Bioinformatics* **21**, 2329-2335.

Hernandez, D., FranÃ§ois, P., Farinelli, L., Ã˜sterÃ¥s, M. & Schrenzel, J. (2008). De novo bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. *Genome Research* **18**, 802-809.

Hintao, J., Teanpaisan, R., Chongsuvivatwong, V., Ratarasan, C. & Dahlen, G. (2007). The microbiological profiles of saliva, supragingival and subgingival plaque and dental caries in adults with and without type 2 diabetes mellitus. *Oral Microbiol Immunol* **22**, 175-181.

Hojo, K., Nagaoka, S., Ohshima, T. & Maeda, N. (2009). Bacterial interactions in dental biofilm development. *J Dent Res* **88**, 982-990. doi: 910.1177/0022034509346811.

Holmberg, K. & Forsum, U. (1973). Identification of Actinomyces, Arachnia, Bacterionema, Rothia, and Propionibacterium species by defined immunofluorescence. *Appl Microbiol* **25**, 834-843.

Homer, K. A. & Beighton, D. (1992). Synergistic degradation of bovine serum albumin by mutans streptococci and other dental plaque bacteria. *FEMS Microbiol Lett* **69**, 259-262.

Hoshino, T., Fujiwara, T. & Kilian, M. (2005). Use of phylogenetic and phenotypic analyses to identify nonhemolytic streptococci isolated from bacteremic patients. *J Clin Microbiol* **43**, 6073-6085.

Howell, A., Jr., Jordan, H. V., Georg, L. K. & Pine, L. (1965). Odontomyces viscosus, gen. nov., spec. nov., a filamentous microorganism isolated from periodontal plaque in hamsters. *Sabouraudia* **4**, 65-68.

Hugenholtz, P., Goebel, B. M. & Pace, N. R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180**, 4765-4774.

Huss, V. A. R., Festl, H. & Schleifer, K. H. (1983). Studies on the spectrophotometric determination of DNA hybridization from renaturation rates. *Systematic and Applied Microbiology* **4**, 184-192.

Ito, T., Katayama, T., Hattie, M. & other authors (2013). Crystal structures of a glycoside hydrolase family 20 lacto-N-biosidase from Bifidobacterium bifidum. *J Biol Chem* **288**, 11795-11806.

Jakubovics, N. S. & Kolenbrander, P. E. (2010). The road to ruin: the formation of disease-associated oral biofilms. *Oral Diseases* **16**, 729-739.

Jensen, P. H., Kolarich, D. & Packer, N. H. (2010). Mucin-type O-glycosylation--putting the pieces together. *Febs J* **277**, 81-94.

Johansson, M. E. V., HolmÃ©n Larsson, J. M. & Hansson, G. C. (2010). The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of hostâ€“microbial interactions. *Proceedings of the National Academy of Sciences*.

Johnson, J. L., Moore, L. V. H., Kaneko, B. & Moore, W. E. C. (1990). Actinomyces georgiae sp. nov., Actinomyces gerencseriae sp. nov., Designation of Two Genospecies of Actinomyces naeslundii, and Inclusion of A. naeslundii serotypes II and III and Actinomyces viscosus serotype II in A. naeslundii Genospecies 2. *International Journal of Systematic Bacteriology* **40**, 273-286.

Jones, N., Bohnsack, J. F., Takahashi, S. & other authors (2003). Multilocus sequence typing system for group B streptococcus. *J Clin Microbiol* **41**, 2530-2536.

Jumas-Bilak, E., Carlier, J. P., Jean-Pierre, H., Teyssier, C., Gay, B., Campos, J. & Marchandin, H. (2004). Veillonella montpellierensis sp. nov., a novel, anaerobic, Gram-negative coccus isolated from human clinical samples. *Int J Syst Evol Microbiol* **54**, 1311-1316.

Jung, I. Y., Choi, B. K., Kum, K. Y., Roh, B. D., Lee, S. J., Lee, C. Y. & Park, D. S. (2000). Molecular epidemiology and association of putative pathogens in root canal infection. *J Endod* **26**, 599-604.

Kamma, J. J., Diamanti-Kipioti, A., Nakou, M. & Mitsis, F. J. (2000). Profile of subgingival microbiota in children with mixed dentition. *Oral Microbiol Immunol* **15**, 103-111.

Katayama, T., Fujita, K. & Yamamoto, K. (2005). Novel bifidobacterial glycosidases acting on sugar chains of mucin glycoproteins. *J Biosci Bioeng* **99**, 457-465.

Kawamura, Y., Hou, X.-G., Sultana, F., Liu, S., Yamamoto, H. & Ezaki, T. (1995). Transfer of Streptococcus adjacens and Streptococcus defectivus to Abiotrophia gen. nov. as Abiotrophia adiacens comb. nov. and Abiotrophia defectiva comb. nov., Respectively. *International Journal of Systematic Bacteriology* **45**, 798-803.

Keene, H. J., Coykendall, A. L. & Lahmeyer, H. A. (1966). Oral glucose clearance time in caries-active and caries-resistant naval recruits. *J Dent Res* **45**, 409.

Keene, H. J., Coykendall, A. L. & Lahmeyer, H. A. (1996). Oral glucose clearance time in caries-active and caries-resistant naval recruits. *J Dent Res* **45**, 409.

Kielty, P. (2011). Oral biology: molecular techniques and applications. *Br Dent J* **211**, 47-47.

Kim, Y. R., Lee, S. E., Kim, C. M. & other authors (2003). Characterization and pathogenic significance of Vibrio vulnificus antigens preferentially expressed in septicemic patients. *Infect Immun* **71**, 5461-5471.

King, S. J., Leigh, J. A., Heath, P. J., Luque, I., Tarradas, C., Dowson, C. G. & Whatmore, A. M. (2002). Development of a multilocus sequence typing scheme for the pig pathogen Streptococcus suis: identification of virulent clones and potential capsular serotype exchange. *J Clin Microbiol* **40**, 3671-3680.

Kiratisin, P., Li, L., Murray, P. R. & Fischer, S. H. (2005). Use of housekeeping gene sequencing for species identification of viridans streptococci. *Diagn Microbiol Infect Dis* **51**, 297-301.

Kitaoka, M., Tian, J. & Nishimoto, M. (2005). Novel putative galactose operon involving lacto-N-biose phosphorylase in Bifidobacterium longum. *Appl Environ Microbiol* **71**, 3158-3162.

Kitaoka, M. (2012). Bifidobacterial enzymes involved in the metabolism of human milk oligosaccharides. *Adv* **3**, 422S-429S. doi: 410.3945/an.3111.001420.

Kiyohara, M., Tachizawa, A., Nishimoto, M., Kitaoka, M., Ashida, H. & Yamamoto, K. (2009). Prebiotic effect of lacto-N-biose I on bifidobacterial growth. *Biosci Biotechnol Biochem* **73**, 1175-1179. Epub 2009 May 1177.

Klein, A., Carnoy, C., Wieruszeski, J. M., Strecker, G., Strang, A. M., van Halbeek, H., Roussel, P. & Lamblin, G. (1992). The broad diversity of neutral and sialylated oligosaccharides derived from human salivary mucins. *Biochemistry* **31**, 6152-6165.

Koehler, A., Karch, H., Beikler, T., Flemmig, T. F., Suerbaum, S. & Schmidt, H. (2003). Multilocus sequence analysis of Porphyromonas gingivalis indicates frequent recombination. *Microbiology* **149**, 2407-2415.

Koeppel, A. F. & Wu, M. (2013). Surprisingly extensive mixed phylogenetic and ecological signals among bacterial Operational Taxonomic Units. *Nucleic Acids Research*.

Kolenbrander, P. E., Inouye, Y. & Holdeman, L. V. (1983). New Actinomyces and Streptococcus coaggregation groups among human oral isolates from the same site. *Infect Immun* **41**, 501-506.

Kolenbrander, P. E. & Andersen, R. N. (1986). Multigeneric aggregations among oral bacteria: a network of independent cell-to-cell interactions. *J Bacteriol* **168**, 851-859.

Kolenbrander, P. E. & Andersen, R. N. (1989). Inhibition of coaggregation between Fusobacterium nucleatum and Porphyromonas (Bacteroides) gingivalis by lactose and related sugars. *Infect Immun* **57**, 3204-3209.

Kolenbrander, P. E., Ganeshkumar, N., Cassels, F. J. & Hughes, C. V. (1993). Coaggregation: specific adherence among human oral plaque bacteria. *Faseb J* **7**, 406-413.

Kolenbrander, P. E. & London, J. (1993). Adhere today, here tomorrow: oral bacterial adherence. *J Bacteriol* **175**, 3247-3252.

Kolenbrander, P. E., Andersen, R. N., Blehert, D. S., Egland, P. G., Foster, J. S. & Palmer, R. J. (2002). Communication among Oral Bacteria. *Microbiology and Molecular Biology Reviews* **66**, 486-505.

Kolenbrander, P. E., Palmer, R. J., Jr., Rickard, A. H., Jakubovics, N. S., Chalmers, N. I. & Diaz, P. I. (2006). Bacterial interactions and successions during plaque development. *Periodontol 2000* **42**, 47-79.

Konstantinidis, K. T. & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* **102**, 2567-2572. Epub 2005 Feb 2568.

Kozitskaya, S., Olson, M. E., Fey, P. D., Witte, W., Ohlsen, K. & Ziebuhr, W. (2005). Clonal analysis of Staphylococcus epidermidis isolates carrying or lacking biofilm-mediating genes by multilocus sequence typing. *J Clin Microbiol* **43**, 4751-4757.

Kriz, P., Kalmusova, J. & Felsberg, J. (2002). Multilocus sequence typing of Neisseria meningitidis directly from cerebrospinal fluid. *Epidemiol Infect* **128**, 157-160.

Kroes, I., Lepp, P. W. & Relman, D. A. (1999). Bacterial diversity within the human subgingival crevice. *Proc Natl Acad Sci U S A* **96**, 14547-14552.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**, 150-163.

Kunz, C. & Rudloff, S. (1993). Biological functions of oligosaccharides in human milk. *Acta Pædiatrica* **82**, 903-912.

Kunz, C., Rudloff, S., Baier, W., Klein, N. & Strobel, S. (2000). Oligosaccharides in human milk: structural, functional, and metabolic aspects. *Annu Rev Nutr* **20**, 699-722.

Kuramitsu, H. K., He, X., Lux, R., Anderson, M. H. & Shi, W. (2007). Interspecies Interactions within Oral Microbial Communities. *Microbiology and Molecular Biology Reviews* **71**, 653-670.

Kuvatanasuchati, J., Chamroensaksri, N. & Tanasupawat, S. (2012). Phenotypic and genotypic characterization of Thai oral streptococci, lactobacilli and pediococci. *Trop* **29**, 254-264.

Lederberg, M. A. T. (2001). Ome sweet omics: a genealogical treasury of words. *Scientist* **15**, 8-10.

Lehmann, P. (1999). P.R. Murray, E.J. Baron, M.A. Pfaller, F.C. Tenover and R.H. Yolken, eds. Manual of Clinical Microbiology, 7th ed. *Mycopathologia* **146**, 107-108.

Leloir, L. F. (1951). The enzymatic transformation of uridine diphosphate glucose into a galactose derivative. *Archives of Biochemistry and Biophysics* **33**, 186-190.

Lendenmann, U., Grogan, J. & Oppenheim, F. G. (2000). Saliva and Dental Pellicle-A Review. *Advances in Dental Research* **14**, 22-28.

Li, J., Helmerhorst, E. J., Leone, C. W., Troxler, R. F., Yaskell, T., Haffajee, A. D., Socransky, S. S. & Oppenheim, F. G. (2004). Identification of early microbial colonizers in human dental biofilm. *J Appl Microbiol* **97**, 1311-1318.

Lie, T. (1977). Early dental plaque morphogenesis. A scanning electron microscope study using the hydroxyapatite splint model and a low-sucrose diet. *J Periodontal Res* **12**, 73-89.

Lie, T. (1979). Morphologic studies on dental plaque formation. *Acta Odontol Scand* **37**, 73-85.

Liebana, J., Parejo, E., Castillo, A. & Gutierrez, J. (1993). Phenotypic characterization of oral streptococci by classical methods. *Microbios* **76**, 7-18.

Lloyd, A. L., Rasko, D. A. & Mobley, H. L. (2007). Defining genomic islands and uropathogen-specific genes in uropathogenic Escherichia coli. *J Bacteriol* **189**, 3532-3546.

LoCascio, R. G., Ninonuevo, M. R., Freeman, S. L., Sela, D. A., Grimm, R., Lebrilla, C. B., Mills, D. A. & German, J. B. (2007). Glycoprofiling of bifidobacterial consumption of human milk oligosaccharides demonstrates strain specific, preferential consumption of small chain glycans secreted in early human lactation. *J Agric Food Chem* **55**, 8914-8919. Epub 2007 Oct 8915.

Lodders, N., Stackebrandt, E. & Nubel, U. (2005). Frequent genetic recombination in natural populations of the marine cyanobacterium Microcoleus chthonoplastes. *Environ Microbiol* **7**, 434-442.

Loesche, W. J. (1986). Role of Streptococcus mutans in human dental decay. *Microbiol Rev* **50**, 353-380.

Lucas, V. S., Beighton, D., Roberts, G. J. & Challacombe, S. J. (1997). Changes in the oral streptococcal flora of children undergoing allogeneic bone marrow transplantation. *J Infect* **35**, 135-141.

Mager, D. L., Ximenez-Fyvie, L. A., Haffajee, A. D. & Socransky, S. S. (2003). Distribution of selected bacterial species on intraoral surfaces. *Journal of Clinical Periodontology* **30**, 644-654.

Mahlen, S. D. & Clarridge, J. E., 3rd (2009). Site and clinical significance of Alloscardovia omnicolens and Bifidobacterium species isolated in the clinical laboratory. *J Clin Microbiol* **47**, 3289-3293.

Maiden, M. C., Bygraves, J. A., Feil, E. & other authors (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**, 3140-3145.

Maiden, M. C. (2006). Multilocus sequence typing of bacteria. *Annu Rev Microbiol* **60**, 561-588.

Maione, D., Margarit, I., Rinaudo, C. D. & other authors (2005). Immunology: Identification of a universal group B Streptococcus vaccine by multiple genome screen. *Science* **309**, 148-150.

Marcotte, H. & Lavoie, M. C. (1998). Oral Microbial Ecology and the Role of Salivary Immunoglobulin A. *Microbiology and Molecular Biology Reviews* **62**, 71-109.

Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* **9**, 387-402.

Marguerat, S., Wilhelm, B.T., and Bahler, J. (2008). Next-generation sequencing: applications beyond genomes. *Biochem Soc Trans* **36**, 1091-1096.

Margulies, M., Egholm, M., Altman, W. E. & other authors (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380. Epub 2005 Jul 2031.

Marsh, P., Martin, M. V. & Lewis, M. A. O. (2009). *Oral Microbiology*: Elsevier.

Marsh, P. D. & Percival, R. S. (2006). The oral microflora--friend or foe? Can we decide? *Int Dent J* **56**, 233-239.

Marsh, P. D., Moter, A. & Devine, D. A. (2010). Dental plaque biofilms: communities, conflict and control. *Periodontology 2000* **55**, 16-35.

Masson, N., Domingues, R. R., Cury, J. A. & Paes Leme, A. F. (2013). Acidulated Phosphate Fluoride Application Changes the Protein Composition of Human Acquired Enamel Pellicle. *Caries Research* **47**, 251-258.

McDowell, A., Perry, A. L., Lambert, P. A. & Patrick, S. (2008). A new phylogenetic group of Propionibacterium acnes. *Journal of Medical Microbiology* **57**, 218-224.

McIntire, F. C., Vatter, A. E., Baros, J. & Arnold, J. (1978). Mechanism of coaggregation between Actinomyces viscosus T14V and Streptococcus sanguis 34. *Infection and Immunity* **21**, 978-988.

Medini, D., Donati, C., Tettelin, H., Masignani, V. & Rappuoli, R. (2005). The microbial pan-genome. *Curr Opin Genet Dev* **15**, 589-594. Epub 2005 Sep 2026.

Meier-Kolthoff, J. P., Goker, M., Sproer, C. & Klenk, H. P. (2013). When should a DDH experiment be mandatory in microbial taxonomy? *Arch Microbiol* **195**, 413-418.

Merritt, J., Tsang, P., Zheng, L., Shi, W. & Qi, F. (2007). Construction of a counterselection-based in-frame deletion system for genetic studies of Streptococcus mutans. *Oral Microbiology and Immunology* **22**, 95-102.

Mestecky, J. (1993). Saliva as a manifestation of the common mucosal immune system. *Ann N Y Acad Sci* **694**, 184-194.

Miller, W. G., On, S. L., Wang, G., Fontanoz, S., Lastovica, A. J. & Mandrell, R. E. (2005). Extended multilocus sequence typing system for Campylobacter coli, C. lari, C. upsaliensis, and C. helveticus. *J Clin Microbiol* **43**, 2315-2329.

Mishra, A., Das, A., Cisar, J. O. & Ton-That, H. (2007). Sortase-catalyzed assembly of distinct heteromeric fimbriae in Actinomyces naeslundii. *J Bacteriol* **189**, 3156-3165.

Mishra, A., Wu, C., Yang, J., Cisar, J. O., Das, A. & Ton-That, H. (2010). The Actinomyces oris type 2 fimbrial shaft FimA mediates co-aggregation with oral streptococci, adherence to red blood cells and biofilm development. *Molecular Microbiology* **77**, 841-854.

Mitra, R. D. & Church, G. M. (1999). In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Research* **27**, e34-e39.

Moncla, B. J. & Braham, P. (1989). Detection of sialidase (neuraminidase) activity in Actinomyces species by using 2'-(4-methylumbelliferyl)alpha-D-N-acetylneuraminic acid in a filter paper spot test. *Journal of Clinical Microbiology* **27**, 182-184.

Munson, M. A., Pitt-Ford, T., Chong, B., Weightman, A. & Wade, W. G. (2002). Molecular and cultural analysis of the microflora associated with endodontic infections. *J Dent Res* **81**, 761-766.

Murray, P. A., Prakobphol, A., Lee, T., Hoover, C. I. & Fisher, S. J. (1992). Adherence of oral streptococci to salivary glycoproteins. *Infection and Immunity* **60**, 31-38.

Mushegian, A. R. & Koonin, E. V. (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci U S A* **93**, 10268-10273.

Nakajima, M., Nihira, T., Nishimoto, M. & Kitaoka, M. (2008). Identification of galacto-N-biose phosphorylase from Clostridium perfringens ATCC13124. *Appl Microbiol Biotechnol* **78**, 465-471. doi: 410.1007/s00253-00007-01319-00258. Epub 02008 Jan 00259.

Nakamura, Y., Itoh, T., Matsuda, H. & Gojobori, T. (2004). Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet* **36**, 760-766.

Naser, S. M., Dawyndt, P., Hoste, B., Gevers, D., Vandemeulebroecke, K., Cleenwerck, I., Vancanneyt, M. & Swings, J. (2007). Identification of lactobacilli by pheS and rpoA gene sequence analyses. *Int J Syst Evol Microbiol* **57**, 2777-2789.

Nasidze, I., Li, J., Quinque, D., Tang, K. & Stoneking, M. (2009). Global diversity in the human salivary microbiome. *Genome Res* **19**, 636-643. doi: 610.1101/gr.084616.084108. Epub 082009 Feb 084627.

Nesbo, C. L., Dlutek, M. & Doolittle, W. F. (2006). Recombination in Thermotoga: implications for species concepts and biogeography. *Genetics* **172**, 759-769. Epub 2005 Dec 2001.

Niazi, S. A., Clarke, D., Do, T., Gilbert, S. C., Mannocci, F. & Beighton, D. (2010). Propionibacterium acnes and Staphylococcus epidermidis Isolated from Refractory Endodontic Lesions Are Opportunistic Pathogens. *Journal of Clinical Microbiology* **48**, 3859-3869.

Nield-Gehrig, W. D. (2003). Dental plaque biofilms. Foundations of Periodontics for the Dental Hygienist. *Philadelphia: Lippincott Williams & Wilkins*, 67-73.

Nishimoto, M. & Kitaoka, M. (2007a). Practical preparation of lacto-N-biose I, a candidate for the bifidus factor in human milk. *Biosci Biotechnol Biochem* **71**, 2101-2104. Epub 2007 Aug 2107.

Nishimoto, M. & Kitaoka, M. (2007b). Identification of N-acetylhexosamine 1-kinase in the complete lacto-N-biose I/galacto-N-biose metabolic pathway in Bifidobacterium longum. *Appl Environ Microbiol* **73**, 6444-6449. Epub 2007 Aug 6424.

Nobbs, A. H., Jenkinson, H. F. & Jakubovics, N. S. (2011). Stick to your gums: mechanisms of oral microbial adherence. *J Dent Res* **90**, 1271-1278.

Nossa, C. W., Oberdorf, W. E., Yang, L. & other authors (2010). Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. *World J Gastroenterol* **16**, 4135-4144.

Nyvad, B. & Kilian, M. (1987). Microbiology of the early colonization of human enamel and root surfaces in vivo. *Scand J Dent Res* **95**, 369-380.

Nyvad, B. & Kilian, M. (1990). Comparison of the initial streptococcal microflora on dental enamel in caries-active and in caries-inactive individuals. *Caries Res* **24**, 267-272.

Nyvad, B., Crielaard, W., Mira, A., Takahashi, N. & Beighton, D. (2013). Dental caries from a molecular microbiological perspective. *Caries Res* **47**, 89-102.

Okahashi, N., Nakata, M., Terao, Y. & other authors (2011). Pili of oral Streptococcus sanguinis bind to salivary amylase and promote the biofilm formation. *Microb Pathog* **50**, 148-154.

Paddick, J. S., Brailsford, S. R., Kidd, E. A., Gilbert, S. C., Clark, D. T., Alam, S., Killick, Z. J. & Beighton, D. (2003). Effect of the environment on genotypic diversity of Actinomyces naeslundii and Streptococcus oralis in the oral biofilm. *Appl Environ Microbiol* **69**, 6475-6480.

Pallen, M. J. & Wren, B. W. (2007). Bacterial pathogenomics. *Nature* **449**, 835-842.

Palmer, R., Jr., Darveau, R., Lamont, R., Nyvad, B. & Teles, R. (2011). Human Oral Bacterial Biofilms: Composition, Dynamics, and Pathogenesis. In *Biofilm Infections*, pp. 35-68: Springer New York.

Palmer, R. J., Diaz, P. I. & Kolenbrander, P. E. (2006). Rapid Succession within the Veillonella Population of a Developing Human Oral Biofilm In Situ. *Journal of Bacteriology* **188**, 4117-4124.

Palys, T., Nakamura, L. K. & Cohan, F. M. (1997). Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int J Syst Bacteriol* **47**, 1145-1156.

Palys, T., Berger, E., Mitrica, I., Nakamura, L. K. & Cohan, F. M. (2000). Protein-coding genes as molecular markers for ecologically distinct populations: the case of two Bacillus species. *Int J Syst Evol Microbiol* **50**, 1021-1028.

Papaioannou, W., Gizani, S., Haffajee, A. D., Quirynen, M., Mamai-Homata, E. & Papagiannoulis, L. (2009). The microbiota on different oral surfaces in healthy children. *Oral Microbiol Immunol* **24**, 183-189.

Paster, B. J., Boches, S. K., Galvin, J. L., Ericson, R. E., Lau, C. N., Levanos, V. A., Sahasrabudhe, A. & Dewhirst, F. E. (2001). Bacterial diversity in human subgingival plaque. *J Bacteriol* **183**, 3770-3783.

Paster, B. J., Olsen, I., Aas, J. A. & Dewhirst, F. E. (2006). The breadth of bacterial diversity in the human periodontal pocket and other oral sites. *Periodontol 2000* **42**, 80-87.

Patton, S. (1994). Detection of Large Fragments of the Human Milk Mucin MUC-1 in Feces of Breast-Fed Infants. *Journal of Pediatric Gastroenterology and Nutrition* **18**, 225-230.

Pedros-Alio, C. (2006). Marine microbial diversity: can it be determined? *Trends Microbiol* **14**, 257-263. Epub 2006 May 2006.

Penders, J., Thijs, C., Vink, C., Stelma, F. F., Snijders, B., Kummeling, I., van den Brandt, P. A. & Stobberingh, E. E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* **118**, 511-521.

Periasamy, S., Chalmers, N. I., Du-Thumm, L. & Kolenbrander, P. E. (2009). Fusobacterium nucleatum ATCC 10953 requires Actinomyces naeslundii ATCC 43146 for growth on saliva in a three-species community that includes Streptococcus oralis 34. *Appl Environ Microbiol* **75**, 3250-3257.

Periasamy, S. & Kolenbrander, P. E. (2010). Central Role of the Early Colonizer Veillonella sp. in Establishing Multispecies Biofilm Communities with Initial, Middle, and Late Colonizers of Enamel. *Journal of Bacteriology* **192**, 2965-2972.

Polonovski, M. & Montreuil, J. (1954). Chromatographic study of the polyosides of human milk. *Comptes rendus hebdomadaires des seances de l'Academie des sciences* **238**, 2263-2264.

Poyart, C., Quesne, G., Coulon, S., Berche, P. & Trieu-Cuot, P. (1998). Identification of streptococci to species level by sequencing the gene encoding the manganese-dependent superoxide dismutase. *J Clin Microbiol* **36**, 41-47.

Preza, D., Olsen, I., Aas, J. A., Willumsen, T., Grinde, B. & Paster, B. J. (2008). Bacterial profiles of root caries in elderly patients. *J Clin Microbiol* **46**, 2015-2021.

Preza, D., Olsen, I., Willumsen, T., Boches, S. K., Cotton, S. L., Grinde, B. & Paster, B. J. (2009). Microarray analysis of the microflora of root caries in elderly. *Eur J Clin Microbiol Infect Dis* **28**, 509-517.

Pullinger, G. D., Lopez-Benavides, M., Coffey, T. J., Williamson, J. H., Cursons, R. T., Summers, E., Lacy-Hulbert, J., Maiden, M. C. & Leigh, J. A. (2006). Application of Streptococcus uberis multilocus sequence typing: analysis of the population structure detected among environmental and bovine isolates from New Zealand and the United Kingdom. *Appl Environ Microbiol* **72**, 1429-1436.

Putnins, E. E. & Bowden, G. H. (1993). Antigenic relationships among oral Actinomyces isolates, Actinomyces naeslundii genospecies 1 and 2, Actinomyces howellii, Actinomyces denticolens, and Actinomyces slackii. *J Dent Res* **72**, 1374-1385.

Ramos, C. P., Foster, G. & Collins, M. D. (1997). Phylogenetic analysis of the genus Actinomyces based on 16S rRNA gene sequences: description of Arcanobacterium phocae sp. nov., Arcanobacterium bernardiae comb. nov., and Arcanobacterium pyogenes comb. nov. *Int J Syst Bacteriol* **47**, 46-53.

Randa, M. A., Polz, M. F. & Lim, E. (2004). Effects of temperature and salinity on Vibrio vulnificus population dynamics as assessed by quantitative PCR. *Appl Environ Microbiol* **70**, 5469-5476.

Reyrat, J. M., Pelicic, V., Gicquel, B. & Rappuoli, R. (1998). Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. *Infect Immun* **66**, 4011-4017.

Rickard, A. H., Gilbert, P., High, N. J., Kolenbrander, P. E. & Handley, P. S. (2003). Bacterial coaggregation: an integral process in the development of multi-species biofilms. *Trends in Microbiology* **11**, 94-100.

Ritz, H. L. (1967). Microbial population shifts in developing human dental plaque. *Arch Oral Biol* **12**, 1561-1568.

Roberts, A. (2005). Bacteria in the mouth. *Dent Update* **32**, 134-136, 139-140, 142.

Robinson, D. A. & Enright, M. C. (2004). Multilocus sequence typing and the evolution of methicillin-resistant Staphylococcus aureus. *Clin Microbiol Infect* **10**, 92-97.

Rolph, H. J., Lennon, A., Riggio, M. P., Saunders, W. P., MacKenzie, D., Coldero, L. & Bagg, J. (2001). Molecular identification of microorganisms from endodontic infections. *J Clin Microbiol* **39**, 3282-3289.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. & Nyren, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* **242**, 84-89.

Rosan, B. & Lamont, R. J. (2000). Dental plaque formation. *Microbes and Infection* **2**, 1599-1607.

Rose, M. C. & Voynow, J. A. (2006). Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiol Rev* **86**, 245-278.

Rossello-Mora, R. & Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiol Rev* **25**, 39-67.

Rotimi, V. O. & Duerden, B. I. (1981). The development of the bacterial flora in normal neonates. *J Med Microbiol* **14**, 51-62.

Ruby, J. D., Li, Y., Luo, Y. & Caufield, P. W. (2002). Genetic characterization of the oral Actinomyces. *Arch Oral Biol* **47**, 457-463.

Ruhl, S., Sandberg, A. L., Cole, M. F. & Cisar, J. O. (1996). Recognition of immunoglobulin A1 by oral actinomyces and streptococcal lectins. *Infection and Immunity* **64**, 5421-5424.

Ruhl, S., Cisar, J. O. & Sandberg, A. L. (2000). Identification of polymorphonuclear leukocyte and HL-60 cell receptors for adhesins of Streptococcus gordonii and Actinomyces naeslundii. *Infect Immun* **68**, 6346-6354.

Sachdeo, A., Haffajee, A. D. & Socransky, S. S. (2008). Biofilms in the edentulous oral cavity. *J Prosthodont* **17**, 348-356.

Sails, A. D., Swaminathan, B. & Fields, P. I. (2003). Clonal complexes of Campylobacter jejuni identified by multilocus sequence typing correlate with strain associations identified by multilocus enzyme electrophoresis. *J Clin Microbiol* **41**, 4058-4067.

Sakamoto, M., Umeda, M., Ishikawa, I. & Benno, Y. (2000). Comparison of the oral bacterial flora in saliva from a healthy subject and two periodontitis patients by sequence analysis of 16S rDNA libraries. *Microbiol Immunol* **44**, 643-652.

Sanders, W., and Sanders, CC. (1984). Modification of normal flora by antibiotics: effects on individuals and the environment. *In: Koot RK, Sande MA, eds* **New Dimensions in Antimicrobial Chemotherapy. New York: Churchill Livingstone**, 217– 241.

Sano, M., Hayakawa, K. & Kato, I. (1992). An enzyme releasing lacto-N-biose from oligosaccharides. *Proc Natl Acad Sci U S A* **89**, 8512-8516.

Sansone, C., Van Houte, J., Joshipura, K., Kent, R. & Margolis, H. C. (1993). The association of mutans streptococci and non-mutans streptococci capable of acidogenesis at a low pH with dental caries on enamel and root surfaces. *J Dent Res* **72**, 508-516.

Sarkonen, N., Kononen, E., Summanen, P., Kanervo, A., Takala, A. & Jousimies-Somer, H. (2000). Oral colonization with Actinomyces species in infants by two years of age. *J Dent Res* **79**, 864-867.

Sarkonen, N., KÃ¶nÃ¶nen, E., Summanen, P., KÃ¶nÃ¶nen, M. & Jousimies-Somer, H. (2001). Phenotypic Identification of Actinomycesand Related Species Isolated from Human Sources. *Journal of Clinical Microbiology* **39**, 3955-3961.

Sato, T., Matsuyama, J., Takahashi, N., Sato, M., Johnson, J., Schachtele, C. & Hoshino, E. (1998). Differentiation of oral Actinomyces species by 16S ribosomal DNA polymerase chain reaction-restriction fragment length polymorphism. *Arch Oral Biol* **43**, 247-252.

Satoh, T., Odamaki, T., Namura, M., Shimizu, T., Iwatsuki, K., Nishimoto, M., Kitaoka, M. & Xiao, J. Z. (2013). In vitro comparative evaluation of the impact of lacto-N-biose I, a major building block of human milk oligosaccharides, on the fecal microbiota of infants. *Anaerobe* **19**, 50-57.

Scannapieco, F. A., Bergey, E. J., Reddy, B. M. S. & Levine, M. J. (1989). Characterization of salivary Î±-amylase binding to Streptococcus sanguis. *Infection and Immunity* **57**, 2853-2863.

Schaal, K., Yassin, A. & Stackebrandt, E. (2006). The Family Actinomycetaceae: The Genera Actinomyces, Actinobaculum, Arcanobacterium, Varibaculum, and Mobiluncus. In *The Prokaryotes*, pp. 430-537. Edited by M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt: Springer New York.

Schaal, K. P. (1986). Genus Actinomyces. In Bergey's manual of systematic bacteriology,. *Edited by PHA Sneath, NSMair, ME Sharpe & J G, Holt Baltimore (MD): Williams and Wilkins*, 1383-1418.

Schaal, K. P. & Lee, H.-J. (1992). Actinomycete infections in humans -- a review. *Gene* **115**, 201-211.

Schaal, K. P. G., R (1985). Serological and numerical phenetic classification of clinically significant fermentative actinomycetes. I. *In Filamentous Microorganisms Biomedical Aspects pp* 85-109.

Schofield, G. M. & Schaal, K. P. (1981). A numerical taxonomic study of members of the Actinomycetaceae and related taxa. *J Gen Microbiol* **127**, 237-259.

Schouls, L. M., Reulen, S., Duim, B., Wagenaar, J. A., Willems, R. J., Dingle, K. E., Colles, F. M. & Van Embden, J. D. (2003). Comparative genotyping of Campylobacter jejuni by amplified fragment length polymorphism, multilocus sequence typing, and short repeat sequencing: strain diversity, host range, and recombination. *J Clin Microbiol* **41**, 15-26.

Schroten, H., Lethen, A., Hanisch, F. G., Plogmann, R., Hacker, J., Nobis-Bosch, R. & Wahn, V. (1992). Inhibition of adhesion of S-fimbriated Escherichia coli to epithelial cells by meconium and feces of breast-fed and formula-fed newborns: mucins are the major inhibitory component. *J Pediatr Gastroenterol Nutr* **15**, 150-158.

Sedgley, C. M., Lee, E. H., Martin, M. J. & Flannagan, S. E. (2008). Antibiotic Resistance Gene Transfer between Streptococcus gordonii and Enterococcus faecalis in Root Canals of Teeth Ex Vivo. *Journal of Endodontics* **34**, 570-574.

Segata, N., Haake, S. K., Mannon, P., Lemon, K. P., Waldron, L., Gevers, D., Huttenhower, C. & Izard, J. (2012). Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol* **13**, 2012-2013.

Sela, D. A., Chapman, J., Adeuya, A. & other authors (2008). The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. *Proc Natl Acad Sci U S A* **105**, 18964-18969. doi: 18910.11073/pnas.0809584105. Epub 0809582008 Nov 0809584124.

Selander, R. K. & Levin, B. R. (1980). Genetic diversity and structure in Escherichia coli populations. *Science* **210**, 545-547.

Shah, D. S. & Russell, R. R. (2004). A novel glucan-binding protein with lipase activity from the oral pathogen Streptococcus mutans. *Microbiology* **150**, 1947-1956.

Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135-1145.

Siqueira, W. L., Custodio, W. & McDonald, E. E. (2012). New Insights into the Composition and Functions of the Acquired Enamel Pellicle. *Journal of Dental Research* **91**, 1110-1118.

Smith, J. M., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993). How clonal are bacteria? *Proc Natl Acad Sci U S A* **90**, 4384-4388.

Smith, K. & Beighton, D. (1986). The effects of the availability of diet on the levels of exoglycosidases in the supragingival plaque of macaque monkeys. *J Dent Res* **65**, 1349-1352.

Smokvina, T., Wels, M., Polka, J., Chervaux, C., Brisse, S., Boekhorst, J., Vlieg, J. E. & Siezen, R. J. (2013). Lactobacillus paracasei Comparative Genomics: Towards Species Pan-Genome Definition and Exploitation of Diversity. *PLoS One* **8**, 22.

Snyder, L. A., Loman, N., Pallen, M.J., and Penn, C.W. (2009). Next-Generation Sequencing-the Promise and Perils of Charting the Great Microbial Unknown. *Microb Ecol* **57**, 1-3.

Spratt, B. G. & Maiden, M. C. (1999). Bacterial population genetics, evolution and epidemiology. *Philos Trans R Soc Lond B Biol Sci* **354**, 701-710.

Spratt, D. A., Weightman, A. J. & Wade, W. G. (1999). Diversity of oral asaccharolytic Eubacterium species in periodontitis--identification of novel phylotypes representing uncultivated taxa. *Oral Microbiol Immunol* **14**, 56-59.

Stackebrandt, E. & Charfreitag, O. (1990). Partial 16S rRNA primary structure of five Actinomyces species: phylogenetic implications and development of an Actinomyces israelii-specific oligonucleotide probe. *J Gen Microbiol* **136**, 37-43.

Stackebrandt, E. & Liesack, W. (1993). Nucleic acids and classification. *In Handbook of New bacterial Systematics: Edited by M Goodfellow & AG O'Donnell London: Academic press*, 151-194.

Stackebrandt, E. & Goebel, B. M. (1994). Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic Bacteriology* **44**, 846-849.

Stackebrandt, E., Frederiksen, W., Garrity, G. M. & other authors (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology* **52**, 1043-1047.

Stackebrandt, E. (2003). The richness of prokaryotic diversity: there must be a species somewhere. *Food Technol Biotechnol* **41**, 17-22.

Stackebrandt, E. & Rossello-Mora, R. (2006). DNA-DNA Reassociation Methods Applied to Microbial Taxonomy and Their Critical Evaluation. In *Molecular Identification, Systematics, and Population Structure of Prokaryotes*, pp. 23-50: Springer Berlin Heidelberg.

Staley, J. T. (2006). The bacterial species dilemma and the genomicâ€“phylogenetic species concept. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**, 1899-1909.

Stoodley, P., Sauer, K., Davies, D. G. & Costerton, J. W. (2002). Biofilms as complex differentiated communities. *Annu Rev Microbiol* **56**, 187-209.

Stramberg, N. & Karlsson, K. A. (1990). Characterization of the binding of Actinomyces naeslundii (ATCC 12104) and Actinomyces viscosus (ATCC 19246) to glycosphingolipids, using a solid-phase overlay approach. *Journal of Biological Chemistry* **265**, 11251-11258.

Stromberg, N. & Karlsson, K. A. (1990). Characterization of the binding of Actinomyces naeslundii (ATCC 12104) and Actinomyces viscosus (ATCC 19246) to glycosphingolipids, using a solid-phase overlay approach. *Journal of Biological Chemistry* **265**, 11251-11258.

Takahashi, N. & Yamada, T. (1999). Glucose and Lactate Metabolism By Actinomyces Naeslundii. *Critical Reviews in Oral Biology & Medicine* **10**, 487-503.

Takahashi, N. (2005). Microbial ecosystem in the oral cavity: Metabolic diversity in an ecological niche and its relationship with oral diseases. *International Congress Series* **1284**, 103-112.

Takahashi, N. & Nyvad, B. (2008). Caries Ecology Revisited: Microbial Dynamics and the Caries Process. *Caries Research* **42**, 409-418.

Tang, G., Yip, H. K., Luo, G., Cheung, B. P., Shen, S. & Samaranayake, L. P. (2003). Development of novel oligonucleotide probes for seven Actinomyces species and their utility in supragingival plaque analysis. *Oral Dis* **9**, 203-209.

Tanner, A., Maiden, M. F. J., Paster, B. J. & Dewhirst, F. E. (1994). The impact of 16S ribosomal RNA-based phylogeny on the taxonomy of oral bacteria. *Periodontology 2000* **5**, 26-51.

Tettelin, H., Masignani, V., Cieslewicz, M. J. & other authors (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A* **102**, 13950-13955. Epub 12005 Sep 13919.

Thomas, C. M. & Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* **3**, 711-721.

Thompson, J. D., Higgins, D.G. & Gibson, T.J. (1994). Clustal-W-Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, Position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680.

Thompson, L. & Lovestedt, S. A. (1951). An Actinomyces-like organism obtained from the human mouth. *Proc Staff Meet Mayo Clin* **26**, 169-175.

Turroni, F., Bottacini, F., Foroni, E. & other authors (2010). Genome analysis of Bifidobacterium bifidum PRL2010 reveals metabolic pathways for host-derived glycan foraging. *Proceedings of the National Academy of Sciences* **107**, 19514-19519.

Urashima, T., Asakuma, S., Leo, F., Fukuda, K., Messer, M. & Oftedal, O. T. (2012). The Predominance of Type I Oligosaccharides Is a Feature Specific to Human Breast Milk. *Advances in Nutrition: An International Review Journal* **3**, 473S-482S.

van der Hoeven, J. S. & Camp, P. J. (1994). The use of lectins in monitoring degradation of oligosaccharide chains in mucin by oral streptococci. *Caries Res* **28**, 257-261.

van Houte, J., Lopman, J. & Kent, R. (1996). The final pH of bacteria comprising the predominant flora on sound and carious human root and enamel surfaces. *J Dent Res* **75**, 1008-1014.

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K. & Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* **60**, 407-438.

Vartoukian, S. R., Palmer, R. M. & Wade, W. G. (2010). Strategies for culture of 'unculturable' bacteria. *FEMS Microbiol Lett* **309**, 1-7.

Venter, J. C., Remington, K., Heidelberg, J. F. & other authors (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science* **304**, 66-74.

Ventura, M., van Sinderen, D., Fitzgerald, G. F. & Zink, R. (2004). Insights into the taxonomy, genetics and physiology of bifidobacteria. *Antonie van Leeuwenhoek* **86**, 205-223.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F. & van Sinderen, D. (2007). Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* **71**, 495-548.

Vos, M. & Didelot, X. (2009). A comparison of homologous recombination rates in bacteria and archaea. *Isme J* **3**, 199-208.

Wada, J., Suzuki, R., Fushinobu, S. & other authors (2007). Purification, crystallization and preliminary X-ray analysis of the galacto-N-biose-/lacto-N-biose I-binding protein (GL-BP) of the ABC transporter from Bifidobacterium longum JCM1217. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **63**, 751-753. Epub 2007 Aug 2010.

Wada, J., Ando, T., Kiyohara, M., Ashida, H., Kitaoka, M., Yamaguchi, M., Kumagai, H., Katayama, T. & Yamamoto, K. (2008). Bifidobacterium bifidum lacto-N-biosidase, a critical enzyme for the degradation of human milk oligosaccharides with a type 1 structure. *Appl Environ Microbiol* **74**, 3996-4004. doi: 3910.1128/AEM.00149-00108. Epub 02008 May 00149.

Wade, W. G., Spratt, D. A., Dymock, D. & Weightman, A. J. (1997). Molecular detection of novel anaerobic species in dentoalveolar abscesses. *Clin Infect Dis* **25**, S235-236.

Ward, R. E., Ninonuevo, M., Mills, D. A., Lebrilla, C. B. & German, J. B. (2006). In vitro fermentation of breast milk oligosaccharides by Bifidobacterium infantis and Lactobacillus gasseri. *Appl Environ Microbiol* **72**, 4497-4499.

Warming, S. r., Costantino, N., Court, D. L., Jenkins, N. A. & Copeland, N. G. (2005). Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Research* **33**, e36.

Wayne, L. G., Brenner, D. J., Colwell, R. R. & other authors (1987). Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *International Journal of Systematic Bacteriology* **37**, 463-464.

Welch, R. A., Burland, V., Plunkett, G., 3rd & other authors (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. *Proc Natl Acad Sci U S A* **99**, 17020-17024.

Wertz, J. E., Goldstone, C., Gordon, D. M. & Riley, M. A. (2003). A molecular phylogeny of enteric bacteria and implications for a bacterial species concept. *J Evol Biol* **16**, 1236-1248.

Whiley, R. (1987). A taxonomic study of oral streptococci. PhD thesis. *London university*.

Whiley, R. A., Fraser, H., Hardie, J. M. & Beighton, D. (1990). Phenotypic differentiation of Streptococcus intermedius, Streptococcus constellatus, and Streptococcus anginosus strains within the "Streptococcus milleri group". *J Clin Microbiol* **28**, 1497-1501.

Whiley, R. A. & Beighton, D. (1998). Current classification of the oral streptococci. *Oral Microbiol Immunol* **13**, 195-216.

Whittaker, C. J., Klier, C. M. & Kolenbrander, P. E. (1996). Mechanisms of adhesion by oral bacteria. *Annual Review of Microbiology* **50**, 513-552.

Wiater, A., Pleszczynska, M., Prochniak, K. & Szczodrak, J. (2012). Structural diversity of streptococcal mutans synthesized under different culture and environmental conditions and its effect on mutanase synthesis. *Molecules* **17**, 11800-11815. doi: 11810.13390/molecules171011800.

Wickstrom, C., Herzberg, M. C., Beighton, D. & Svensater, G. (2009). Proteolytic degradation of human salivary MUC5B by dental biofilms. *Microbiology* **155**, 2866-2872. Epub 2009 Jun 2825.

Wilson, M. & Wilson, M. (2004). *Role of the indigenous microbiota in maintaining human health*
Microbial Inhabitants of Humans: Cambridge University Press.

Wilson, M. (2005). Microbial inhabitants of humans. Their ecology and role in health and disease.

Winslow, C. E., Broadhurst, J., Buchanan, R. E., Krumwiede, C., Rogers, L. A. & Smith, G. H. (1920). The Families and Genera of the Bacteria: Final Report of the Committee of the Society of American Bacteriologists on Characterization and Classification of Bacterial Types. *J Bacteriol* **5**, 191-229.

Wirth, T., Falush, D., Lan, R. & other authors (2006). Sex and virulence in Escherichia coli: an evolutionary perspective. *Mol Microbiol* **60**, 1136-1151.

Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev* **51**, 221-271.

Wu, C. & Ton-That, H. (2010). Allelic exchange in Actinomyces oris with mCherry fluorescence counterselection. *Appl Environ Microbiol* **76**, 5987-5989.

Ximenez-Fyvie, L. A., Haffajee, A. D., Martin, L., Tanner, A., Macuch, P. & Socransky, S. S. (1999). Identification of oral Actinomyces species using DNA probes. *Oral Microbiol Immunol* **14**, 257-265.

Ximénez-Fyvie, L. A., Haffajee, A. D. & Socransky, S. S. (2000). Microbial composition of supra- and subgingival plaque in subjects with adult periodontitis. *Journal of Clinical Periodontology* **27**, 722-732.

Yeung, M. K. & Ragsdale, P. A. (1997). Synthesis and function of Actinomyces naeslundii T14V type 1 fimbriae require the expression of additional fimbria-associated genes. *Infection and Immunity* **65**, 2629-2639.

Yeung, M. K., Donkersloot, J. A., Cisar, J. O. & Ragsdale, P. A. (1998). Identification of a Gene Involved in Assembly ofActinomyces naeslundii T14V Type 2 Fimbriae. *Infection and Immunity* **66**, 1482-1491.

Yolken, R. H., Peterson, J. A., Vonderfecht, S. L., Fouts, E. T., Midthun, K. & Newburg, D. S. (1992). Human milk mucin inhibits rotavirus replication and prevents experimental gastroenteritis. *The Journal of Clinical Investigation* **90**, 1984-1991.

Yoshida, E., Sakurama, H., Kiyohara, M. & other authors (2012). Bifidobacterium longum subsp. infantis uses two different beta-galactosidases for selectively degrading type-1 and type-2 human milk oligosaccharides. *Glycobiology* **22**, 361-368.

Yoshioka, H., Iseki, K. & Fujita, K. (1983). Development and differences of intestinal flora in the neonatal period in breast-fed and bottle-fed infants. *Pediatrics* **72**, 317-321.

Zee, K. Y., Samaranayake, L. P. & Attstrom, R. (2001). Salivary immunoglobulin A levels in rapid and slow plaque formers: a pilot study. *Microbios* **106**, 81-87.

Zuckerkandl, E. & Pauling, L. (1965). Molecules as documents of evolutionary history. *Journal of Theoretical Biology* **8**, 357-366.